Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 6   ISSN 1438-8871   Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Short Paper

Letters to the Editor

Corrigenda and Addendas

Review

Adolescent Peer Influence on Eating Behaviors via Social Media: Scoping Review

Alicia Chung[1], EdD; Dorice Vieira[2], MPH; Tiffany Donley[2], DSC; Nicholas Tan[3], BA; Girardin Jean-Louis[2], PhD; Kathleen Kiely Gouley[1], PhD; Azizi Seixas[2], PhD

[1]Center for Early Childhood Health and Development, Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States

[2]NYU Grossman School of Medicine, New York, NY, United States

[3]SUNY Downstate College of Medicine, Brooklyn, NY, United States

Corresponding Author:
Alicia Chung, EdD
Center for Early Childhood Health and Development
Department of Population Health
NYU Grossman School of Medicine
227 East 30th Street
New York, NY
United States
Phone: 1 212 263 1359
Email: alicia.chung@nyumc.org

Abstract

Background: The influence of social media among adolescent peer groups can be a powerful change agent.

Objective: Our scoping review aimed to elucidate the ways in which social media use among adolescent peers influences eating behaviors.

Methods: A scoping review of the literature of articles published from journal inception to 2019 was performed by searching PubMed (ie, MEDLINE), Embase, CINAHL, PsycINFO, Web of Science, and other databases. The review was conducted in three steps: (1) identification of the research question and clarification of criteria using the population, intervention, comparison, and outcome (PICO) framework; (2) selection of articles from the literature using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines; and (3) charting and summarizing information from selected articles. PubMed's Medical Subject Headings (MeSH) and Embase's Emtree subject headings were reviewed along with specific keywords to construct a comprehensive search strategy. Subject headings and keywords were based on adolescent age groups, social media platforms, and eating behaviors. After screening 1387 peer-reviewed articles, 37 articles were assessed for eligibility. Participant age, gender, study location, social media channels utilized, user volume, and content themes related to findings were extracted from the articles.

Results: Six articles met the final inclusion criteria. A final sample size of 1225 adolescents (aged 10 to 19 years) from the United States, the United Kingdom, Sweden, Norway, Denmark, Portugal, Brazil, and Australia were included in controlled and qualitative studies. Instagram and Facebook were among the most popular social media platforms that influenced healthful eating behaviors (ie, fruit and vegetable intake) as well as unhealthful eating behaviors related to fast food advertising. Online forums served as accessible channels for eating disorder relapse prevention among youth. Social media influence converged around four central themes: (1) visual appeal, (2) content dissemination, (3) socialized digital connections, and (4) adolescent marketer influencers.

Conclusions: Adolescent peer influence in social media environments spans the spectrum of healthy eating (ie, pathological) to eating disorders (ie, nonpathological). Strategic network-driven approaches should be considered for engaging adolescents in the promotion of positive dietary behaviors.

(J Med Internet Res 2021;23(6):e19697) doi:10.2196/19697

KEYWORDS

social media; eating behaviors; adolescent health

Introduction

Defining Social Media

Adolescent peer groups have been recognized to influence individuals' health behaviors, including diet [1]. During adolescence, eating behaviors are influenced by peer impacts, such as perceived social norms that can create unique peer pressures [2,3]. Peer-to-peer influence on health behaviors has been documented in face-to-face interactions [4]; however, few have studied the influence of social media on eating behaviors during adolescence.

Social media has been defined as any social networking site that enables interactive, user-generated content that allows sharing of images, ideas, videos, music, or commentary on internet forums (eg, Facebook), blogs and microblogs (eg, Twitter), and photograph- or video-hosting platforms (eg, Instagram, YouTube, or TikTok) [5]. Individuals or groups of people can communicate, collaborate, and connect in real time via text, video, or phone anywhere that Wi-Fi is available. Social media channels, such as Facebook or YouTube, were initiated in the early 2000s. However, the first website recognized as being the first social media platform was called *Six Degrees*—short for Six Degrees of Separation—and it launched in 1997. In 2018, YouTube, Instagram, and Snapchat were identified as the most popular online platforms utilized by teens 13 to 17 years of age [6]. User-generated content on these channels may allow for autonomy, identity, and interpersonal peer relationship development, a hallmark of adolescence [7].

Social media is an effective channel for engaging adolescents [8], a target population that has been hard to engage in public health practice. It can be used to influence, inform, and persuade. Social media mobile apps have global reach, use, and engagement [9]. In an earlier global report, approximately 85% of adolescents between the ages of 12 and 17 years across Europe, Latin America, the United States, and South Korea reported using a social media website [10]. Among a sample of 4460 high school students from Turkey in 2019, 88% owned a smartphone and 100% had a social media account [11]. Contagion effect—the rapid communication of an idea that has gone *viral* among peers on social media platforms—has been recognized as an effective way to promote health behaviors [12-15]. Behavior intent, increased knowledge, and increased awareness are positive attributes of healthful food posts on social media that influence users [16-18]. Extensive social media use, along with other entertainment media use, has been associated with consumption of unhealthy foods, mostly due to snacking behaviors. In particular, Albert found that social media and other entertainment media use among a sample of mostly Latino (68%) middle schoolers was negatively correlated with fruit and vegetable consumption (r=–0.065) and was strongly correlated with fast food and junk food intake (r<0.200) [19]. In a recent report, Chau et al concluded that social media was a promising channel for obesity prevention in adolescents and young adults [20]. Given that more recent research revealed that 95% of teens 13 to 17 years of age own a smartphone, 51% use Facebook, 69% use Snapchat, 72% use Instagram, and 85% use YouTube [21], an examination of peer influence, via social media channels, on eating behaviors is warranted. However, no review to date has demonstrated peer influence on eating behaviors via social media networks among adolescents.

Social Media Influence and Eating Behaviors

A social network analysis of adult, in-person peer relationship influences indicated that maladaptive eating behaviors (ie, eating disorders) may be influenced by friendships [22]. Social norms, as well as real and perceived social support, may be underpinning peer influences related to the practice of eating. Peer groups and the type and degree of peer influence may shape one's relationship with food. Peer influence on eating behaviors may extend from in-person influence to social media influence. Findings from a US nationally representative sample of young adults, 19 to 32 years of age, revealed an association between a high volume and frequency of social media platform engagement (ie, Facebook, Twitter, Google+, YouTube, LinkedIn, Instagram, Pinterest, Tumblr, Vine, Snapchat, and Reddit) and eating concerns [23]. However, some of the most popular social media channels have been noted to influence maladaptive (ie, nonpathological) eating disorders as well as adaptive (ie, pathological) healthy eating.

Social media platforms (ie, Facebook and YouTube) and mobile gaming nutrition-intervention apps (eg, Food Hero) demonstrate utility among young adult populations to raise awareness, increase knowledge, influence intrinsic beliefs, and motivate attitudes [23]. Social media channels, including Facebook, YouTube, and Snapchat, have been recognized by adolescents for providing peer-to-peer support in healthy eating through sharing information and offering social support [24]. This scoping review aimed to elucidate the role of peer influence via social media channels on eating behaviors among adolescents between the ages of 10 and 19 years.

Methods

Databases

The following databases were searched in October 2017 and updated in October 2019: PubMed (ie, MEDLINE), AgeLine, BIOSIS Citation Index, CINAHL, the Cochrane Library, Embase, ERIC (Education Resources Information Center), Food Science and Technology Abstracts, Google Scholar, Inspec, PubMed Central, PsycINFO, SciELO (Scientific Electronic Library Online), and Web of Science. PubMed's Medical Subject Headings (MeSH) and Embase's Emtree subject headings were reviewed along with specific keywords to construct a comprehensive search strategy. Grey literature was searched for in The New York Academy of Medicine Grey Literature Report and the OAIster database from the OCLC (Online Computer Library Center). An extensive electronic journal hand search was conducted in the following journals: American Journal of Health Promotion, American Journal of Preventive Medicine, Appetite, Childhood Obesity, Eating Behavior, Ethnicity & Disease, Ethnicity & Health, International Journal of Eating Disorders, International Journal of Obesity (London), Journal of the American Dietetic Association, Pediatrics, Obesity (Silver Spring), and Public Health Nutrition. In consultation with the first author (AC), a clinical librarian (DV) trained in systematic literature reviews conducted the literature search and managed the information tools. The project was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [25]. A comprehensive search strategy of subject headings and keywords included “obesity,” “nutrition,” “peer behavior,” and “adolescents” (see Multimedia Appendix 1 for full set of terms).

Inclusion and Exclusion Criteria

Inclusion criteria for articles to be included in the review were as follows: study sample included adolescents 10 to 19 years of age; study examined a social media app; study had a cross-sectional, qualitative, observational, and experimental design; study had a social media component; study examined adolescent peer communications in a social media environment; and study examined eating behaviors. Studies were required to be written in English and were published from journal inception to October 2019. Conversely, studies related to the impact of social media on body image, related to gastric bypass, or conducted in animal models were excluded. EndNote X9 (Clarivate Analytics) was used to manage the bibliographic data. All references were downloaded to Google Sheets for screening. Full texts were retrieved and a Google Form was created for data extraction.

Data Extraction

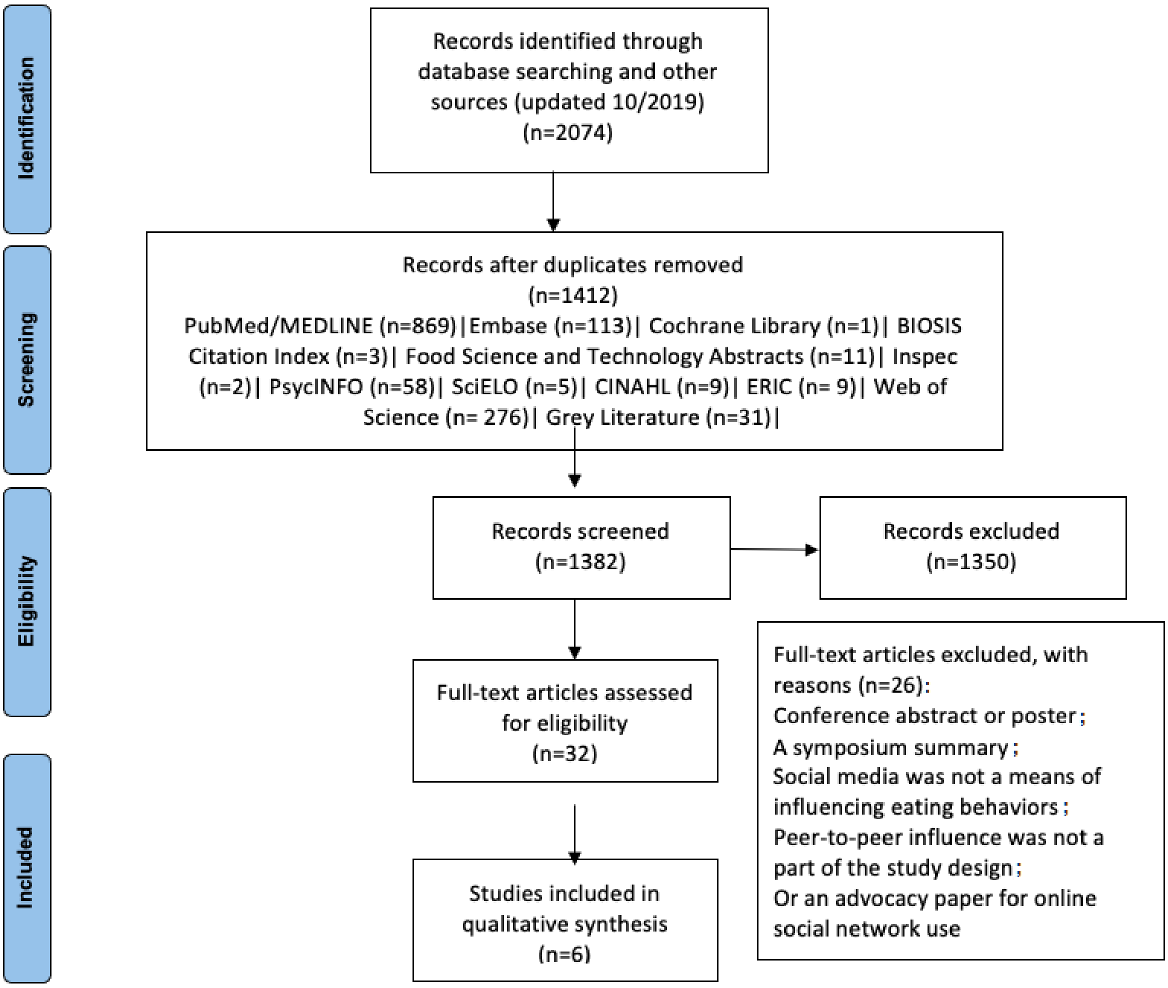
Two independent researchers (NT and AS) screened the articles, assessed them for eligibility, and extracted the data from the search results. The extracted data were exported to Microsoft Excel 2016 for data analysis. Specifically, extracted data included author names; year of publication; country of study; study time frame; participant ages and genders; total number of participants; total number of user accounts; racial and ethnic groups, including percentage or whole number by group; study design; type of social media; behavior influence on food; and primary study outcomes measured. Reviewer agreement was reached through discussion with the senior author and review of the abstracts. Reviews of the study titles, abstracts, and full text, where needed, were completed to ensure agreement with study inclusion parameters to confirm eligibility. Tiebreakers were decided by DV.

Results

Overview

Figure 1 shows the PRISMA flowchart for the article selection process for this review.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for the articles that met the final inclusion process. Grey literature includes The New York Academy of Medicine Grey Literature Report and the OAIster database from the OCLC (Online Computer Library Center). ERIC: Education Resources Information Center; SciELO: Scientific Electronic Library Online.



A total of 1225 adolescents ranging in age from 10 to 19 years who participated in studies across the United States, the United Kingdom, Sweden, Norway, Denmark, Portugal, Brazil, and Australia were represented in the sample. Social media platforms included Facebook and Facebook Messenger, YouTube, Instagram, Twitter, self-made blogs, school websites, and researcher-moderated forums. Each study examined how youth utilized such social media platforms to communicate ideas regarding food and diet.

A total of 2074 articles were identified through electronic databases and manual hand searching of articles from systematic review reference lists. After removing duplicates, 1412 articles remained and were evaluated according to our inclusion criteria. Of the 32 full-text articles that were reviewed further, 26 did not meet the inclusion criteria and were removed. A total of six studies met the full inclusion criteria and were included in the final sample (Figure 1). Table 1 shows the details of these six studies [26-31].

Instagram and Facebook were among the most popular social media platforms that influenced healthful eating behaviors (ie, fruit and vegetable intake) as well as unhealthful eating behaviors related to fast food advertising. Online forums served as accessible channels for eating disorder relapse prevention among youth. However, self-made blogs on anorexia also promoted content about self-harming behavior in support of the eating disorder.

Both positive and negative influences were found in eating behavior content among adolescents. Holmberg et al [26,27] were the only researchers to leverage a contemporary social media platform (ie, Instagram) and identify positive eating behaviors promoted among adolescents. Their study setting across Sweden, Norway, and Denmark is a leading example of the utility of social media to influence eating behaviors in a positive way. Their engagement strategy with adolescents could be leveraged for future studies that evaluate actual eating behavior change. Food safety behaviors were the other positive aspect of healthy eating promoted via social media in the United States, as identified by Quick et al [28]. Kendal et al [29] described self-made blogs for relapse prevention of eating disorders. Fast food advertising, as identified by Thaichon and Quach [30], and self-harming anorexic eating behavior content, as identified by Castro and Osório [31], had negative influences on adolescent eating behaviors (Table 1).

Thematically, results of the six studies included in the scoping review revealed several core themes related to engagement and dissemination of food-related content in a social media environment among adolescent users across eight developed countries. Social media influence converged around four central themes: (1) visual appeal, (2) content dissemination, (3) socialized digital connections, and (4) adolescent *marketer* influencers. Social media not only served as a communication channel to a target cohort but also allowed for interaction through a platform that allowed for user autonomy on a specific topic.

Table 1. Data extraction of key variables from each of the six studies that met the final inclusion criteria.

Authors	Age group (years)	Country of study setting	Social media engagement and theme	Main outcomes	Participants
Castro and Osório [31]	13-19	Portugal and Brazil	11 Portuguese-language blogs in Portugal and Brazil Three categories analyzed: 1. Self-harming content 2. Celebrities and fashion models as body image "thinspiration" 3. Proanorexic testimonials about dealing with parental, peer, and other social and cultural offline pressure	The internet is a powerful means of supporting the *proanorexic* movement. This contributes to consumption and production of problematic blog content.	2 boys and 9 girls
Quick et al [28]	11-13	United States	Facebook, YouTube, Oovvuu, and Skype Theme: Food safety promotion videos to improve behaviors among American middle schoolers were disseminated via peer social media networks	Viewing the videos increased perceived susceptibility of food-borne illness and increased self-confidence in performing food safety behaviors.	21 boys and 23 girls
Kendal et al [29]	10-19	United Kingdom	Engagement: 420 message postings 119 usernames; 97 threads Platform: one self-made online forum Themes: 1. Mentorship 2. Online forum as a safe space 3. Friendship within the online forum 4. Flexible help 5. Peer support for recovery and relapse prevention	Online discussion forum was used to help manage eating disorders and overcome maladaptive eating behaviors.	119 unique usernames
Holmberg et al [26]	14	Sweden, Norway, and Denmark	Engagement: 3479 Instagram images 1712 available accounts 1001 Instagram accounts with the hashtag #14 (pertaining to age groups)	Food images were found in most adolescent accounts. Food was often centrally placed and framed in a positive way. Images of food with high calories and low nutrients were framed as advertisements. Images depicting fruits and vegetables were often zoomed in on, similar to images found in cookbooks.	1001 unique Instagram accounts
Thaichon and Quach [30]	11-16	Australia	Platforms: Facebook and Twitter Themes: 1. Peer pressure: users try to match their page through like and share functions via social networks 2. Eating habits and intentions: children said that they tend to change their eating habits after repeatedly being exposed to advertisements on social networking sites	Online marketing via social media sites had a strong impact on children's decisions to consume fast food.	15 boys and 15 girls
Holmberg et al [27]	13-16	Sweden	Platform: Instagram Engagement: semistructured interviews that described adolescent social media engagement with food Themes: 1. Protecting self-esteem by not disclosing body weight, body images, or unhealthy foods, to minimize the risk of receiving hurtful comments 2. Props and symbols that had positive associations among their peers, which signified social status, that could generate likes and positive comments were favored	Depictions of body image and food self-presentation in digital social media were the most prominent influential imagery.	11 girls and 9 boys

Theme 1: Influence of Visually Appealing Imagery and Endorsements

First, communicating eye-catching visual imagery of food, fitness, and body ideals among adolescent peers played a primary role of influence in several modalities. Videos and pictures allow for creative design feature elements that may increase engagement. Entertaining videos with appealing graphics, relatable scenarios, and music were employed effectively in Quick et al’s development of food safety videos. These videos increased food safety practices (ie, handwashing) among 332 preadolescent peer groups in an experimental design study within the United States [28]. Similarly, Holmberg and colleagues [26] reported that shared images of food in social media may reflect a lifestyle that adolescents admire or want to promote. Positive framing of fruits and vegetables that were colorful and aesthetically pleasing may be indicative of a certain status worth sharing [26]. Sharing food images or videos that were perceived as preferable by peers reflect endorsements and may encourage the likelihood of behavior adoption.

Similarly, *likes* or other affirmative visual icons, such as hearts and smiley emojis, connote positive agreement and affirm approval of what is depicted by a statement or image. Food brands with positive associations among peers that signified social status and generated *likes* and positive comments were viewed as favorable among teens in Sweden. In a qualitative study with Swedish teens, one 14-year-old participant described her reasoning for sharing a Starbucks image as follows: “Even if one has never had a Starbucks beverage or visited the place, one still loves it, because one knows that everyone else loves it” [27].

Peer group reinforcers may influence in both directions of the pendulum. Findings of semistructured interviews with 20 Swedish adolescent boys and girls enrolled in a pediatric obesity clinic revealed that they avoided posting unhealthy “fattening” foods in fear of this behavior being viewed as unacceptable from their peers with the potential to elicit criticism and bullying [27]. Posting visual images of healthy foods was used to portray acceptability, as these foods would be viewed favorably and positively by peers. Conversely, Castro and Osório [31] reported that attractive imagery may also influence aspirational ideals toward thinness—or “thinspiration”—among teen bloggers in Portugal and Brazil on proanorexic websites [31].

Theme 2: Social Media Dissemination

Second, posting images and videos with friends allows for quick dissemination of health-related ideas, products, and practices. Social media has the ability to be circulated worldwide instantly. Rapid dissemination allows for contagion *virality* or *viral marketing* of a topic faster than a formal broadcasting channel, indicating the speed of dissemination by sharing unique, entertaining messages on one’s pre-existing social network [28]. Quick communication uptakes may bypass mainstream media channel dissemination and speed.

The emotion conveyed behind *viral* messages may influence person-to-person well-being. Social media dissemination of harmful ideas may have negative consequences on adolescent mental health, thereby influencing nonpathological eating disorders. Social media contagion passed between individuals may have mental health implications contributing to unique social norms that affect anxiety levels due to *viral* messaging of negative eating behaviors.

Theme 3: Socializing Digital Connections

Third, digital platforms facilitate peer-to-peer interactions in ways that are quicker and more convenient than traditional in-person support networks, potentially influencing peer social norms. Social media interactions occur at a speed, engagement, and influence level beyond in-person communication. Their reach and scale could effectively influence adolescent beliefs, attitudes, and norms around eating behaviors at a broad scale. Digital communities have been leveraged by the food industry as a marketing tool to advertise directly to consumers. Thaichon and Quach [30] reported that food advertising to young consumers associated a company’s product with community and socialization. The authors found that fast food advertising on social media influenced adolescent views toward fast food, eating habits, and purchasing likelihood [30].

Socialization of digital connections builds online community relationships that connect individuals based on shared experiences, including eating behaviors. Social media digital communities have had positive and negative effects on eating disorder behaviors among adolescents. Castro and Osório [31] were able to engage adolescents in real time across 11 Portuguese-language blogs in Portugal and Brazil in challenges with anorexia through shared cultural pressures and struggles of living with an eating disorder [31]. Similarly, Kendal and colleagues [29] were able to garner peer support in the form of proactive self-care for relapse prevention of anorexic eating behaviors through online forums among adolescents in the United Kingdom [29]. Online access any time of day served as an accessible resource for peer support that allowed for flexible support, friendship in a “safe environment,” and peer support for relapse prevention [29].

Theme 4: Adolescent Influencer Marketer

Images shared by adolescents with one another may be more influential than commercial advertising. Adolescent user-generated food content on social media was presented and received differently than food advertisements, but still mimicked that of food advertisers. Holmberg et al [26] examined how teenagers presented the food they posted online, analyzing trends in food and drink items and how they were described. The authors found most of the food images (68%) depicted high-calorie, nutrient-poor foods, and only 22% of images included fruits and vegetables. Topic engagement allows for peers to relate to each other through common interest and language. Online relationship building is fostered regardless of differences, such as weight status, which could be rendered absent in a digital world of user self-generated content. In comparison, Thaichon and Quach [30] analyzed how the presence of fast food advertisements on social media influenced the dietary opinions of the adolescents who viewed them. Their exploratory qualitative data found that peer communication on social media was a highly influential factor on purchasing behavior, attitudes toward fast food, and eating behaviors. Table

2 [26-31] shows the influences and eating behavior outcomes related to various social media channels.

Thematic review of the selected articles also revealed a mixed pattern of effect of peer influence on eating behavior of adolescents in social media environments. Social media channels were found to be used as social support for both positive and negative eating patterns. Peer-to-peer support for overweight or obese female adolescents who used Facebook Messenger increased their positive perception about social support and, thus, their online social interactions compared to a group that only received large-group face-to-face support [32]. Positive outcomes were also noted in a moderated online forum for adolescents with eating disorders [29]. The digital modality offered assistance in ways that more traditional services could not, such as by enhancing choice, privacy, and control. Conversely, self-harming proanorexic online content was found within a small group of adolescent blogs in Portugal and Brazil. Peer pressure, need for acceptance, and conflicts with parents were social and cultural pressures that youth were grappling with online [31].

Table 2. Influences and eating behavior outcomes by social media channel.

Social media channel	Influences and/or outcomes
Facebook and Twitter	Fast food advertisements can influence young children by the promotion of fast food products and complimentary toys [30].
Instagram	Adolescents presented food images with lifestyle depiction intention in mind to their peers. Positive connotation of fruits and vegetable posts were found. Conclusions were limited due to images not fully representing daily eating. Outcome: food items presented in adolescent social media content and how they were presented were measured [26]. How food items were presented (ie, still-life photos): 20% of food items were arranged as an exhibition, 37.2% were branded food images, and 74.8% included positive adjectives and symbols. What items were presented (ie, types of food): 67.7% of images contained high-calorie, low-nutrient foods and 21.8% contained fruits, vegetables, and berries. Fruit and vegetable images were generally depicted as more visually appealing based on camera zoom and captions. Food items and props were used to protect body image and self-esteem [27].
Various self-made blogs	Reading proanorexia, blogs had no effect on dietary consumption. Content analysis of proanorexia blogs suggests that adolescents pursue harmful minimal food consumption as a result of social and cultural body image, peer pressure and bullying, celebrity and fashion model “thinspiration,” and general exposure to thin-ideal imagery. The blogs themselves contain numerous proanorexia resources and “tips” [31]. This suggests that the internet and social media can serve to promulgate harmful and extreme dietary ideas, although actual behavioral effects were not studied [29].
Online peer-networking eating behavior interventions	Peer-created intervention materials have the potential to reinforce positive nutrition behaviors related to weight loss and food safety among adolescents [28].

Discussion

Principal Findings

The literature on peer-enhanced social media interventions for eating behaviors is in its nascent stages. This scoping review aims to fill the gap in the literature and to review the evidence on the influence of peer-to-peer enhanced social media environments on eating behaviors among adolescent youth aged 10 to 19 years. Self-reported, user-generated eating behavior content on social media, supplanted with image recognition, food diaries, nutrient-intake mobile apps, or data synced to wearable devices, such as cameras embedded in eyeglasses, allows for passive data collection with minimal user burden; this data could be integrated into social media in order to build medical evidence to support decision making. Our paper demonstrates that peer social media influence on dietary behaviors warrants a robust amount of additional work to add to the body of scientific medical evidence in the field.

Holmberg and colleagues [26] reported positive portrayals of healthy eating promoted by adolescents. Fruit and vegetable images that are zoomed-in on and focused on for a picture may place emphasis on the food depicted, due to visual appeal and positive attributes. Poelman et al provide an example of a digital food tracking system that could be embedded into social media apps to understand how food choices are influenced by the real-world food environment [33]. Another option is the use of digital food record mobile apps, such as FitNinja (Vibrent Health), with image recognition software to collect nutrient content; these have been found to be acceptable tools for digital food records of real-world food intake [34]. Additionally, shared food posts, such as fruits and vegetables marked by peer *likes* among user networks in social media environments, may represent reinforcement of positive—or any valence—nutrition behaviors as positive, well-liked behaviors [35].

Commercial advertising on Facebook and Twitter, as described by Thaichon and Quach [30], may detract from adolescent engagement, as teens may seek to declare their independence outside of the mainstream; in addition, these platforms are targeted to older age groups. Social media platforms may allow teens a digital environment for creative license, personal identity, and autonomy during a time frame when they are transcending into early adulthood and away from parental influence [7]. In addition, Instagram and Snapchat, which were launched in 2010 and 2011, respectively, are messaging apps whose early adopters are nearly a generation younger than Facebook users. Facebook may not be as relatable, given its inception with a college cohort in 2004, a generation currently approaching middle age.

Peer influence via social media could be an effective channel to engage this typically hard-to-reach population on health topics, including health behaviors. Social media networks were a consistent setting for engaging adolescents with healthy eating messages [36]. Visual appeal was a strong engagement characteristic that influenced users both positively and negatively. Unfortunately, fast food advertising is also pervasive and influential on social media channels targeting adolescents, which could have negative consequences on weight status and other chronic disease risks [37].

Facebook was the most common social media network reported, despite the rising popularity of Instagram and Snapchat over Facebook among adolescents [6]. Only one Swedish study [26] analyzed adolescents' perception of food on Instagram. This may be due to the time lag in research. One advantage to this could be that as adolescents move away from Facebook, they may be less exposed to the commercial fast food marketing commonly reported on that social media channel.

Healthy eating posts may reflect an aspirational lifestyle change among people in the contemplation phase toward healthy eating. Kinard [38] and Holmberg et al [27] found that obese and overweight adolescents and adults were more likely to engage with healthy food posts than with unhealthy *junk food* posts on Instagram and Facebook [38]. Similarly, Holmberg et al [25] commented that fruits and vegetables were portrayed in a favorable way that connoted palatability. Health promotion marketing of healthy foods may aid to inspire healthful behavior change as users are drawn to the visual appeal.

As social norms are modified in a digital milieu, cautionary monitoring of peer pressures may be needed. Social media peer pressures may affect body image ideals [39] that could lead to maladaptive eating behaviors and poor well-being.

Healthful Social Media Interventions for Adolescent Eating Behaviors

Multipronged interventions with in-person and social media components have reported successful weight loss among participants [40] and an increase in feelings of social support in adolescent populations [32]. Kulik and colleagues [32] reported that social networking builds peer social support for weight loss in conjunction with an in-person intervention. Peer support may offer teens safe space to share emotional vulnerability, where they can relate to and confide in peers, while also serving as a source of accountability for healthful dietary goals. Similarly, significant weight loss was found as a result of a weight loss intervention that used Facebook private messaging and text messaging among a diverse group of college students [41]. Additionally, Barragan et al found that social media platforms (ie, Facebook, Twitter, and YouTube) increased knowledge on excess calorie intake from sugar-sweetened beverages and increased self-reported intention to reduce sugar-sweetened beverage consumption [16]. Additionally, online discussion forums served as a source of mental health support for eating disorder recovery and relapse prevention [29].

Nutrition information may raise awareness and promote nutrition literacy when content is verified. Mixed messages in the media on the healthfulness of certain foods may be misleading to the public [42]. Additionally, dietary information shared on social media is oftentimes misaligned with national dietary guidelines and evidence-based dietary recommendations. Nutrition content on social media needs to be both accurate and engaging to avoid increasing consumer confusion and skepticism of dietary advice altogether [43]. Public health practitioners, nutrition educators, and researchers need to partner with food industry advertisers, social media influencers, and social marketing leaders to ensure that consumers are accurately informed, particularly for vulnerable populations such as adolescents.

Negative Effects of Social Media on Adolescent Eating Behaviors

Social media may influence poor eating habits and maladaptive eating behaviors. Thaichon and Quach [30] reported an association between overweight and obese Australian adolescents and behavior intent toward eating fast food due to advertisements viewed on Facebook. Incentive advertising combined with fast food and soda endorsed by their peers may reinforce the promotion of unhealthy food choices. Additionally, two European studies [29,31] engaged adolescents around maladaptive eating behaviors related to eating disorders. Users provided each other with tips and strategies for bulimic or anorexic eating behaviors, promulgating harmful eating behaviors and extreme diets.

Limitations of the Current Literature

Unfortunately, these studies do not help in understanding the role of social media influence or impact in real-world dietary behavior change in adolescent peer groups. Measurements of actual behavior change need to be studied in conjunction with social media marketing campaigns (eg, purchasing behavior and food intake). Hawkins et al [44] reported that perceived norms and preferences around eating among a sample of English university students (mean age 22 years) on Facebook were predictive of users' actual food consumption [44]. Facebook users' perceived social norms were predictive of users' actual fruit and vegetable intake, and perceived social norms were predictive of participants' actual snack and sugar-sweetened beverage consumption. Also, MySpace and Reddit were not included as social media platforms in the search terms list. Omission of MySpace may account for reduced representation by Black and other racial and ethnic minority groups.

Future Directions

Future research should emphasize methodological rigor to elucidate peer influence on dietary behavior change. An extensive amount of research is needed in the field, including objective measures of actual dietary intake with social media interventions and social network analysis of peer influence change agents on food behavior outcomes. In a pilot study that examined whether promoting red peppers via a social media influencer on Instagram would increase actual vegetable intake among adolescents in the Netherlands, no effect was found on users' actual dietary intake [45]. Additional work is needed to understand the influence of peer-to-peer behavior transmission and adoption in social media environments. The lack of appropriate medical evidence to support decision making might be resolved with more research studies utilizing social media channels alongside objective eating behavior measures. Social media geographic location check-in tools could build off of this approach.

Additionally, fact checking of user-generated content and use of credible dietary sources on social media may be questionable. Content verification of nutrition information [18] may also be affected by perception of friendship ties [46]. Perceived degrees of connection and measurement, or lack thereof, of health outcomes are also limitations when understanding the utility of social media use for adolescent health behaviors [46].

Future research may also include Snapchat and other novel platforms that are now pervasively used by youth [6]. TikTok is also a popular social media platform that was released in 2016 by ByteDance.com that is gaining popularity, particularly during the COVID-19 global pandemic. This video sharing social networking service started in China and gained traction in the United States in 2018 after merging with musical.ly. The social media channel allows users to create short lip sync, dance, and comedic videos [47].

Racial and ethnic youth of color are underrepresented in studies of this kind. Only Kulik et al, who conducted a study in the United States, included minority youth; in their study of Facebook as a complement to an in-person weight loss intervention, 20% of the sample was African American and 21% were participants from other groups of color [32]. Since non-Hispanic Black (22%) and Hispanic (26%) youth experience obesity rates consistently higher than their White counterparts (14%) [48], more research is needed to understand the impact of social media influence on eating behaviors in adolescents of color.

Racial and ethnic health disparities experienced by people of color give rise to a heightened need for targeted healthful marketing via social media channels to engage youth. Racial and ethnic minority youth are heavily targeted for fast food marketing [49,50], and communities of color tend to be inundated by food swamps (ie, an abundance of fast food restaurants concentrated in a ZIP Code). Therefore, in order to act against these high-calorie, nutrient-poor advertising messages [51], culturally tailored approaches are needed to promote healthful eating behaviors among this population [52]. In addition, health literacy has been identified as a key social determinant of health among adolescents [53]. Accurate nutrition-related health literacy conveyed through photos, video imagery, and text is critical to addressing diet-related comorbidities among adolescent youths of color.

Future research should evaluate the role of social media engagement with peer influencer change agents in dietary behavior change interventions. The pervasiveness of social media usage among adolescents calls attention to a communication channel that cannot be ignored. Moreover, the cell phone technology that allows touchscreen access to social media may enhance the capacity of peer influencer change agents that could be more powerful than prior print or television media. In the social media realm, evidence from social network analysis indicates that peer influencers are effective health behavior change agents based on leadership styles by peers, social network connectedness, and communication patterns between the peer influencer change agent and end recipient [54]. Even prior to the global popularity of social media, peer influencers were highly regarded change agents. Peer educator change agents were the most commonly used HIV prevention framework, as peer change agents were more likely to be recognized for their leadership qualities [55]. Gender differences may also be explored in future research about adolescent influence on eating behaviors in social media environments. Constant cell phone engagement offers a technology medium that could not only engage adolescents about eating behaviors but could also support adoption of targeted change behaviors.

Conclusions

Social media offers the potential of a hand-held change agent. Social media use on cell phones has become a global mainstay in contemporary culture, particularly for adolescents. Adolescent youth can serve as digital beacons of influence on health topics, including eating behaviors. Drawing from influencer marketing strategies in the digital landscape, tailored for culture and audience, adolescents could have a significant influence on the health behaviors of their peers. Health promotion initiatives to influence adolescent youth should consider the integration of social media channels.

Acknowledgments

The authors would like to thank Jelyn Eustaquio, Kevin Chery, Rafy Iqbal, and Dishari Azad for their contributions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of the full set of search terms.
[DOCX File , 13 KB - jmir_v23i6e19697_app1.docx]

References

1. Rice EL, Klein WMP. Interactions among perceived norms and attitudes about health-related behaviors in US adolescents. Health Psychol 2019 Mar;38(3):268-275 [FREE Full text] [doi: 10.1037/hea0000722] [Medline: 30762406]

2. Gaspar de Matos M, Palmeira AL, Gaspar T, De Wit JBF, Luszczynska A. Social support influences on eating awareness in children and adolescents: The mediating effect of self-regulatory strategies. Glob Public Health 2016;11(4):437-448. [doi: 10.1080/17441692.2015.1094106] [Medline: 26564992]
3. Sharps M, Robinson E. Perceived eating norms and children's eating behaviour: An informational social influence account. Appetite 2017 Jun 01;113:41-50 [FREE Full text] [doi: 10.1016/j.appet.2017.02.015] [Medline: 28192218]
4. Madan A, Moturu ST, Lazer D, Pentland AS. Social sensing: Obesity, unhealthy eating and exercise in face-to-face networks. In: Proceedings of the Wireless Health Conference. 2010 Presented at: Wireless Health Conference; October 4-7, 2010; San Diego, CA p. 104-110. [doi: 10.1145/1921081.1921094]
5. Ventola C. Social media and health care professionals: Benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
6. Anderson M, Jiang J. Pew Research Center. Washington, DC: Pew Research Center; 2018 May 31. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2021-05-10]
7. Spies Shapiro LA, Margolin G. Growing up wired: Social networking sites and adolescent psychosocial development. Clin Child Fam Psychol Rev 2014 Mar;17(1):1-18 [FREE Full text] [doi: 10.1007/s10567-013-0135-1] [Medline: 23645343]
8. Goodman J, Wennerstrom A, Springgate BF. Participatory and social media to engage youth: From the Obama campaign to public health practice. Ethn Dis 2011;21(3 Suppl 1):S1-S94 [FREE Full text] [Medline: 22352086]
9. Meshi D, Tamir DI, Heekeren HR. The emerging neuroscience of social media. Trends Cogn Sci 2015 Dec;19(12):771-782. [doi: 10.1016/j.tics.2015.09.004] [Medline: 26578288]
10. Hur J, Gupta M. Growing up in the web of social networking: Adolescent development and social media. Adolesc Psychiatry 2013 Aug 01;3(3):233-244. [doi: 10.2174/2210676611303030004]
11. Coskun S, Karayagız Muslu G. Investigation of problematic mobile phones use and fear of missing out (FoMO) level in adolescents. Community Ment Health J 2019 Aug;55(6):1004-1014. [doi: 10.1007/s10597-019-00422-8] [Medline: 31209716]
12. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman Isler M. Social media and HIV: A systematic review of uses of social media in HIV communication. J Med Internet Res 2015 Nov 02;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
13. Yonker LM, Zan S, Scirica CV, Jethwani K, Kinane TB. "Friending" teens: Systematic review of social media in adolescent and young adult health care. J Med Internet Res 2015 Jan 05;17(1):e4 [FREE Full text] [doi: 10.2196/jmir.3692] [Medline: 25560751]
14. Bottorff JL, Struik LL, Bissell LJ, Graham R, Stevens J, Richardson CG. A social media approach to inform youth about breast cancer and smoking: An exploratory descriptive study. Collegian 2014;21(2):159-168. [doi: 10.1016/j.colegn.2014.04.002] [Medline: 25109215]
15. Moreno MA, Whitehill JM. Influence of social media on alcohol use in adolescents and young adults. Alcohol Res 2014;36(1):91-100 [FREE Full text] [Medline: 26259003]
16. Barragan NC, Noller AJ, Robles B, Gase LN, Leighs MS, Bogert S, et al. The "sugar pack" health marketing campaign in Los Angeles County, 2011-2012. Health Promot Pract 2014 Mar;15(2):208-216. [doi: 10.1177/1524839913507280] [Medline: 24149214]
17. Norman CD, Yip AL. eHealth promotion and social innovation with youth: Using social and visual media to engage diverse communities. Stud Health Technol Inform 2012;172:54-70. [Medline: 22910502]
18. Leak TM, Benavente L, Goodell LS, Lassiter A, Jones L, Bowen S. EFNEP graduates' perspectives on social media to supplement nutrition education: Focus group findings from active users. J Nutr Educ Behav 2014;46(3):203-208. [doi: 10.1016/j.jneb.2014.01.006] [Medline: 24613446]
19. Albert SL. #consumingitall: Understanding the Complex Relationship Between Media Consumption and Eating Behaviors [doctoral dissertation]. Los Angeles, CA: University of California, Los Angeles; 2017. URL: https://escholarship.org/uc/item/1f75j34r [accessed 2021-05-10]
20. Chau M, Burgermaster M, Mamykina L. The use of social media in nutrition interventions for adolescents and young adults-A systematic review. Int J Med Inform 2018 Dec;120:77-91 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.10.001] [Medline: 30409348]
21. Wong CA, Merchant RM, Moreno MA. Using social media to engage adolescents and young adults with their health. Healthc (Amst) 2014 Dec;2(4):220-224 [FREE Full text] [doi: 10.1016/j.hjdsi.2014.10.005] [Medline: 25984444]
22. Forney KJ, Schwendler T, Ward RM. Examining similarities in eating pathology, negative affect, and perfectionism among peers: A social network analysis. Appetite 2019 Jun 01;137:236-243 [FREE Full text] [doi: 10.1016/j.appet.2019.03.013] [Medline: 30872142]
23. Sidani JE, Shensa A, Hoffman B, Hanmer J, Primack BA. The association between social media use and eating concerns among US young adults. J Acad Nutr Diet 2016 Sep;116(9):1465-1472 [FREE Full text] [doi: 10.1016/j.jand.2016.03.021] [Medline: 27161027]
24. Klassen KM, Douglass CH, Brennan L, Truby H, Lim MSC. Social media use for nutrition outcomes in young adults: A mixed-methods systematic review. Int J Behav Nutr Phys Act 2018 Jul 24;15(1):70 [FREE Full text] [doi: 10.1186/s12966-018-0696-y] [Medline: 30041699]

25. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche P, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. Ann Intern Med 2009 Aug 18;151(4):W65-W94. [doi: 10.7326/0003-4819-151-4-200908180-00136] [Medline: 19622512]
26. Holmberg C, Chaplin JE, Hillman T, Berg C. Adolescents' presentation of food in social media: An explorative study. Appetite 2016 Apr 01;99:121-129. [doi: 10.1016/j.appet.2016.01.009] [Medline: 26792765]
27. Holmberg C, Berg C, Hillman T, Lissner L, Chaplin JE. Self-presentation in digital media among adolescent patients with obesity: Striving for integrity, risk-reduction, and social recognition. Digit Health 2018;4:1-15 [FREE Full text] [doi: 10.1177/2055207618807603] [Medline: 30349733]
28. Quick V. Short food safety videos promote peer networking and behavior change. Br Food J 2015;117(1):78-93. [doi: 10.1108/bfj-09-2013-0270]
29. Kendal S, Kirk S, Elvey R, Catchpole R, Pryjmachuk S. How a moderated online discussion forum facilitates support for young people with eating disorders. Health Expect 2017 Feb;20(1):98-111 [FREE Full text] [doi: 10.1111/hex.12439] [Medline: 26725547]
30. Thaichon P, Quach TN. Online marketing communications and childhood's intention to consume unhealthy food. Australas Mark J 2021 Jan 27;24(1):79-86. [doi: 10.1016/j.ausmj.2016.01.007]
31. Castro TS, Osório AJ. “I love my bones!”1 – Self-harm and dangerous eating youth behaviours in Portuguese written blogs. Young Consum 2013 Nov 15;14(4):321-330. [doi: 10.1108/yc-03-2013-00351]
32. Kulik NL, Fisher EB, Ward DS, Ennett ST, Bowling JM, Tate DF. Peer support enhanced social support in adolescent females during weight loss. Am J Health Behav 2014 Sep 01;38(5):789-800. [doi: 10.5993/ajhb.38.5.16]
33. Poelman MP, van Lenthe FJ, Scheider S, Kamphuis CB. A smartphone app combining global positioning system data and ecological momentary assessment to track individual food environment exposure, food purchases, and food consumption: Protocol for the observational Foodtrack study. JMIR Res Protoc 2020 Jan 28;9(1):e15283 [FREE Full text] [doi: 10.2196/15283] [Medline: 32012100]
34. Fowler LA, Yingling LR, Brooks AT, Wallen GR, Peters-Lawrence M, McClurkin M, et al. Digital food records in community-based interventions: Mixed-methods pilot study. JMIR Mhealth Uhealth 2018 Jul 17;6(7):e160 [FREE Full text] [doi: 10.2196/mhealth.9729] [Medline: 30021705]
35. Pisano S, Adams M, Koester K, Bruening M. Acceptability of fruit and vegetable marketing materials for schools among adolescents. In: Proceedings of Nutrition 2019. 2019 Jun Presented at: Nutrition 2019; June 8-11, 2019; Baltimore, MD URL: https://www.eventscribe.com/2019/ASN/ajaxcalls/PosterInfo.asp?efp=UURQVExBSVA3OTU5&PosterID=203256&rnd=0.6161305 [doi: 10.1093/cdn/nzz050.p16-043-19]
36. Murphy G, Corcoran C, Tatlow-Golden M, Boyland E, Rooney B. See, like, share, remember: Adolescents' responses to unhealthy-, healthy- and non-food advertising in social media. Int J Environ Res Public Health 2020 Mar 25;17(7):2181 [FREE Full text] [doi: 10.3390/ijerph17072181] [Medline: 32218252]
37. Montgomery KC, Chester J. Interactive food and beverage marketing: Targeting adolescents in the digital age. J Adolesc Health 2009 Sep;45(3 Suppl):S18-S29. [doi: 10.1016/j.jadohealth.2009.04.006] [Medline: 19699433]
38. Kinard BR. Insta-grams: The effect of consumer weight on reactions to healthy food posts. Cyberpsychol Behav Soc Netw 2016 Aug;19(8):481-486. [doi: 10.1089/cyber.2016.0085] [Medline: 27494330]
39. Cohen R, Irwin L, Newton-John T, Slater A. #bodypositivity: A content analysis of body positive accounts on Instagram. Body Image 2019 Jun;29:47-57. [doi: 10.1016/j.bodyim.2019.02.007] [Medline: 30831334]
40. Nour M, Yeung SH, Partridge S, Allman-Farinelli M. A narrative review of social media and game-based nutrition interventions targeted at young adults. J Acad Nutr Diet 2017 May;117(5):735-752.e10. [doi: 10.1016/j.jand.2016.12.014] [Medline: 28238894]
41. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31 [FREE Full text] [doi: 10.1002/oby.20232] [Medline: 23505165]
42. Kininmonth AR, Jamil N, Almatrouk N, Evans CEL. Quality assessment of nutrition coverage in the media: A 6-week survey of five popular UK newspapers. BMJ Open 2017 Dec 27;7(12):e014633 [FREE Full text] [doi: 10.1136/bmjopen-2016-014633] [Medline: 29284712]
43. Ramachandran D, Kite J, Vassallo AJ, Chau JY, Partridge SR, Freeman B, et al. Food trends and popular nutrition advice online - Implications for public health. Online J Public Health Inform 2018;10(2):e213 [FREE Full text] [doi: 10.5210/ojphi.v10i2.9306] [Medline: 30349631]
44. Hawkins LK, Farrow C, Thomas JM. Do perceived norms of social media users' eating habits and preferences predict our own food consumption and BMI? Appetite 2020 Jun 01;149:104611. [doi: 10.1016/j.appet.2020.104611] [Medline: 31958481]
45. Folkvord F, de Bruijne M. The effect of the promotion of vegetables by a social influencer on adolescents' subsequent vegetable intake: A pilot study. Int J Environ Res Public Health 2020 Mar 26;17(7):2243 [FREE Full text] [doi: 10.3390/ijerph17072243] [Medline: 32225032]

46. Almaatouq A, Radaelli L, Pentland A, Shmueli E. Are you your friends' friend? Poor perception of friendship ties limits the ability to promote behavioral change. PLoS One 2016;11(3):e0151588 [FREE Full text] [doi: 10.1371/journal.pone.0151588] [Medline: 27002530]
47. ByteDance. URL: https://www.bytedance.com/en [accessed 2020-04-16]
48. Childhood obesity facts: Prevalence of childhood obesity in the United States. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/obesity/data/childhood.html [accessed 2021-05-10]
49. Lucan SC, Maroko AR, Sanon OC, Schechter CB. Unhealthful food-and-beverage advertising in subway stations: Targeted marketing, vulnerable groups, dietary intake, and poor health. J Urban Health 2017 Apr;94(2):220-232 [FREE Full text] [doi: 10.1007/s11524-016-0127-9] [Medline: 28271237]
50. Herrera AL, Pasch KE. Ethn Health 2018 Aug;23(6):691-702 [FREE Full text] [doi: 10.1080/13557858.2017.1290217] [Medline: 28277028]
51. Cooksey-Stowers K, Schwartz MB, Brownell KD. Food swamps predict obesity rates better than food deserts in the United States. Int J Environ Res Public Health 2017 Nov 14;14(11):1366 [FREE Full text] [doi: 10.3390/ijerph14111366] [Medline: 29135909]
52. Chung A, Wallace B, Stanton-Koko M, Seixas A, Jean-Louis G. Feasibility and acceptability of a culturally tailored website to increase fruit and vegetable intake and physical activity levels in African American mother-child dyads: Observational study. JMIR Pediatr Parent 2019 Mar 22;2(1):e12501 [FREE Full text] [doi: 10.2196/12501] [Medline: 31518320]
53. Fleary SA, Joseph P, Pappagianopoulos JE. Adolescent health literacy and health behaviors: A systematic review. J Adolesc 2018 Jan;62:116-127. [doi: 10.1016/j.adolescence.2017.11.010] [Medline: 29179126]
54. Lee YF, McLaws M, Ong LM, Amir Husin S, Chua HH, Wong SY, et al. Hand hygiene - Social network analysis of peer-identified and management-selected change agents. Antimicrob Resist Infect Control 2019;8:195 [FREE Full text] [doi: 10.1186/s13756-019-0644-x] [Medline: 31798841]
55. Schneider JA, Zhou AN, Laumann EO. A new HIV prevention network approach: Sociometric peer change agent selection. Soc Sci Med 2015 Jan;125:192-202 [FREE Full text] [doi: 10.1016/j.socscimed.2013.12.034] [Medline: 24518188]

Abbreviations

ERIC: Education Resources Information Center
MeSH: Medical Subject Headings
OCLC: Online Computer Library Center
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SciELO: Scientific Electronic Library Online

Edited by T Rashid Soron; submitted 28.04.20; peer-reviewed by L Brennan, L Gardner, MA Bahrami, W Averhart; comments to author 03.06.20; revised version received 30.09.20; accepted 03.03.21; published 03.06.21.

Please cite as:
Chung A, Vieira D, Donley T, Tan N, Jean-Louis G, Kiely Gouley K, Seixas A
Adolescent Peer Influence on Eating Behaviors via Social Media: Scoping Review
J Med Internet Res 2021;23(6):e19697
URL: https://www.jmir.org/2021/6/e19697
doi:10.2196/19697
PMID:34081018

Review

Application of the eHealth Literacy Model in Digital Health Interventions: Scoping Review

Mariam El Benny[1], MPH; Tamar Kabakian-Khasholian[1], PhD; Fadi El-Jardali[2], PhD; Marco Bardus[1], BA, MA, PhD

[1]Department of Health Promotion and Community Health, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon
[2]Department of Health Management and Policy, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

Corresponding Author:
Marco Bardus, BA, MA, PhD
Department of Health Promotion and Community Health
Faculty of Health Sciences
American University of Beirut
PO-Box 11-0236, Riad-El-Solh
Beirut, 1107 2020
Lebanon
Phone: 961 1 350 0000 ext 4694
Email: marco.bardus@gmail.com

Abstract

Background: Digital health interventions (DHIs) are increasingly being adopted globally to address various public health issues. DHIs can be categorized according to four main types of technology: mobile based, web based, telehealth, and electronic health records. In 2006, Norman and Skinner introduced the *eHealth literacy model*, encompassing six domains of skills and abilities (basic, health, information, scientific, media, and computer) needed to effectively understand, process, and act on health-related information. Little is known about whether these domains are assessed or accounted for in DHIs.

Objective: This study aims to explore how DHIs assess and evaluate the eHealth literacy model, describe which health conditions are addressed, and which technologies are used.

Methods: We conducted a scoping review of the literature on DHIs, based on randomized controlled trial design and reporting the assessment of any domain of the eHealth literacy model. MEDLINE, CINAHL, Embase, and Cochrane Library were searched. A duplicate selection and data extraction process was performed; we charted the results according to the country of origin, health condition, technology used, and eHealth literacy domain.

Results: We identified 131 unique DHIs conducted in 26 different countries between 2001 and 2020. Most DHIs were conducted in English-speaking countries (n=81, 61.8%), delivered via the web (n=68, 51.9%), and addressed issues related to noncommunicable diseases (n=57, 43.5%) or mental health (n=26, 19.8%). None of the interventions assessed all six domains of the eHealth literacy model. Most studies focused on the domain of health literacy (n=96, 73.2%), followed by digital (n=19, 14.5%), basic and media (n=4, 3%), and information and scientific literacy (n=1, 0.7%). Of the 131 studies, 7 (5.3%) studies covered both health and digital literacy.

Conclusions: Although many selected DHIs assessed health or digital literacy, no studies comprehensively evaluated all domains of the eHealth literacy model; this evidence might be overlooking important factors that can mediate or moderate the effects of these interventions. Future DHIs should comprehensively assess the eHealth literacy model while developing or evaluating interventions to understand how and why interventions can be effective.

(J Med Internet Res 2021;23(6):e23473) doi:10.2196/23473

KEYWORDS

eHealth literacy; digital health interventions; consumer health information; scoping review; mHealth; mobile phone

Introduction

Digital Health Interventions

In the last 20 years, digital health or eHealth has emerged as an important research field. At the intersection of medical informatics, public health, and business, eHealth refers to the use of "health services and information delivered or enhanced through the Internet and related technologies" [1]. Technologies such as web based, mobile based, telehealth, and electronic health records (EHRs) have become widely adopted in the so-called *digital health interventions* (DHIs). DHIs can be defined as "health services delivered electronically through formal or informal care. DHIs can range from electronic medical records used by providers to mobile health (mHealth) apps used by consumers" [2]. The World Health Organization has recently produced a classification of DHIs, identifying four main types: clients, health care providers, health systems, and data services [3]. On PubMed, as of August 3, 2020, the number of records mentioning *eHealth* or *DHIs* in their title or abstract has consistently increased over the past 20 years, starting from 65 in 2000 to 11,395 in 2019, reaching a total of 6720.

Some systematic reviews and meta-analyses have described the effectiveness of DHIs in addressing various public health problems, such as somatic diseases [4], or health literacy and health outcomes [5]. Nevertheless, it is still unclear what makes DHIs superior to nondigital interventions or what components of these interventions facilitate positive outcomes reported [6]. In addition, it is unclear whether DHIs are effective because of their content or the manner in which they are delivered. Regarding the content of interventions, some systematic reviews have focused on exploring the way people process and understand information available on the internet [7,8]. In fact, with so many resources and information available on the internet, patients and users enrolled in DHIs may face challenges in understanding and making sense of the information they receive. Some research has focused on problems related to the ability to process information derived from web-based sources or delivered through technologies.

The eHealth Literacy Model

In 2006, Norman and Skinner [9] proposed a conceptual model that encompasses six different domains of literacy required to process information from technology sources: traditional literacy, health literacy, information literacy, scientific literacy, media literacy, and computer literacy. According to Norman and Skinner [9], traditional or basic functional literacy includes simple and primitive literacy skills, including the ability to read and understand text and the ability to speak and write in a certain language. Information literacy includes the ability to know how knowledge is structured and how information can be used in a certain way that informs other people. Media literacy is the capability to critique a media subject and place information in different contexts. Health literacy, coined in the 1970s, can be generally defined as "the degree to which individuals can obtain, process, understand, and communicate about health-related information needed to make informed health decisions" (as reported by Berkman et al [10]). According to Norman and Skinner [9], health literacy is the ability to perform basic reading and numerical tasks required to function in the health care environment; patients with adequate health literacy can read, understand, and act on health care information. More recent evolutions of the concept include a variety of competencies and skills, including knowledge, motivation, and competencies related to accessing, understanding, appraising, and applying health-related information in health care, disease prevention, and health promotion settings [11]. Several systematic reviews have analyzed the relationship between health literacy and a variety of health outcomes, indicating that a good level of health literacy is generally associated with positive health outcomes in various health domains, such as vaccination [12], noncommunicable diseases (NCDs) such as chronic kidney disease or coronary artery disease, heart failure [13-15], oral health [16], quality of life [17], and excess body weight [18]. Some other review evidence has shown how interventions promoting critical health literacy [19] could be very beneficial for the community [20] or among specific segments of the population, such as adolescents [21] or older adults [22].

Strictly related to the concept of eHealth is computer or technology literacy, which is the capability to use new technologies and software and the ability to access electronic health information [9]. Recent conceptualizations expand this domain to look at the ability to process information, to engage with patients' own health, at the motivation and ability to engage with digital devices, at feeling safe and in control, at having access to health care and technological systems that work, and at meeting digital services that suit individuals' needs [23]. Norman and Skinner [24] have developed a scale to assess eHealth literacy, called *eHealth literacy scale* (*eHEALS*), which has been one of the most adopted and cited, with 449 citations on the *Journal of Medical Internet Research* page and more than 1320 results on Google Scholar (as of August 3, 2020). The last domain of the eHealth literacy model, scientific literacy, involves the ability to allocate health-related findings in the right context by systematically understanding the "nature, aims, methods, applications, limitations, and politics" of building knowledge [9]. Several systematic reviews have analyzed the relationship between health literacy in mHealth apps and interventions [5,7,8,25,26], generally reporting positive associations among health literacy, digital literacy, and health outcomes. Other reviews have specifically examined how technology can affect health literacy in health programs [27-29].

According to the developers of the eHealth literacy model, the six domains can be grouped into two main categories: analytic (traditional, media, and information) and context-specific (health, scientific, and computer). The analytical category refers to a set of competencies that can be applied to a wide range of information sources, whereas context-specific categories include competencies that can only be applied to a specific problem in a specific context [9]. For example, the ability of a person living with type 2 diabetes to process information related to diabetes is different from their ability to process information related to vaccines, mental health, or other chronic conditions. Similarly, the ability to use a mobile phone to call someone does not necessarily translate into the ability to use a mobile app, navigate a website, or evaluate the information retrieved while searching on the internet.

Related Work and Study Aims

Arguably, researchers developing DHIs should always take into account the domains of computer or technology literacy and health literacy, as these are potential pathways for more effective and equitable interventions [30]. Health literacy can be viewed as both an outcome and a mediator in interventions intended to improve health outcomes [31]. Technologies or delivery modes can also be seen as interacting or moderating factors [32], depending on the type of technology used to deliver an intervention on a specific health topic. DHIs can be developed to improve health literacy (outcome) or they can be developed to improve clinical outcomes in which one or more dimensions of the eHealth literacy model are considered as mediators or moderators of the effects of the intervention. Researchers developing DHIs could then assume that people enrolling in these interventions should have good levels of functional, scientific, media, and information literacy to understand how to write or read information they are exposed to.

However, to what extent are these assumptions tenable? In other words, is the eHealth literacy model purely conceptual or does it find a concrete application in DHIs? To the best of our knowledge, there are no systematic reviews that specifically discuss the application of the complete eHealth literacy model in DHIs. When we were developing the search strategies for this project, we searched for existing systematic reviews in PubMed and PROSPERO databases with the keyword *eHealth literacy* and identified only four systematic reviews [33-36]. However, all these reviews have focused on the domain of digital literacy, looking at specific health outcomes in specific segments of the population, such as people living with HIV [33], underserved populations in the United States [34], older adults [35], or college students [36]. Therefore, this scoping review aims to identify and describe DHIs that assess any domain of the eHealth literacy model and to identify which domains are assessed and evaluated the most. We considered DHIs that were developed to improve clinical outcomes or that were aimed at different literacies, according to the eHealth literacy model. In other words, we considered interventions that looked at eHealth literacy either as an outcome or as a mediator of intervention effects, as long as the domains of the eHealth literacy model were assessed.

Methods

Overview

We followed the scoping review framework by Arksey and O'Malley [37], which encompasses five stages: (1) identification of the initial research questions; (2) identification of relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results. The stages are described further in the following sections.

Stage 1: Identifying the Research Question

The main review question, based on the eHealth literacy model, was "To what extent are DHIs assessing the 6 domains of the eHealth literacy model?" More specifically, we wanted to answer the following research questions: What domains of the eHealth literacy model (ie, computer, health, traditional, media, information, and science literacy) are assessed and reported in the literature? What health conditions have been investigated? What technologies are used?

Stage 2: Identifying Relevant Studies

We searched four electronic databases that cover most of the medical and public health literature: MEDLINE, CINAHL, Embase, and Cochrane Library.

We used a predefined search strategy, encompassing keywords and medical subject headings to cover three main concepts: *eHealth literacy model*, *digital health*, and the *study design* for interventions. The *eHealth literacy model* concept entailed terms such as health literacy, literacy, computer literacy, information literacy, basic, functional, scientific, media, information, computer, health, eHealth, literacy, literate, illiteracy, and illiterate. The second concept, *digital health*, expanded on the above and entailed keywords, such as *telemedicine*, *internet*, *mobile*, *phone*, *digital*, *medium* or *media*, *mHealth*, *eHealth*, *telemedicine*, and *computer*, based on other systematic reviews recently conducted by one of the authors [6,38,39]. The third concept, that is, the research design, entailed a predefined set of keywords and operands that Cochrane has developed to identify randomized controlled trials (RCTs); this is because we wanted to identify the best level of evidence available [40]. The search strategy used for MEDLINE is provided in Multimedia Appendix 1. Database searches were completed, and references were retrieved on January 24, 2020.

In addition, we used the reference list of identified systematic reviews on the topic to identify other potentially relevant studies.

Stage 3: Study Selection

We followed the Joanna Briggs' Institute's PCC (Population-Concept-Context) framework [41,42] to define our inclusion criteria, as it applies to scoping reviews. We included studies that discussed DHIs (concepts) and reported the assessment of at least one domain of the eHealth literacy model (context). In this context, we conceived the dimensions of the eHealth literacy model as either outcomes or mediators of DHIs. The assessment of the different types of literacy was considered a sufficient indicator for DHIs considering such dimensions as outcomes or mediators of intervention effects. We did not restrict the results to any population, with the idea of inductively categorizing the results according to health condition, hence defining the population of reference in the analytical phase.

The screening process consisted of two stages: title and abstract as well as full-text screening. The first stage involved 2 reviewers (MEB and MB) and one research assistant, who independently screened all records identified by the searches. This task was completed using a web-based application for systematic reviews, Rayyan [43]. The interrater reliability was excellent (agreement 96%; Cohen κ=0.834; Gwet AC1=0.950). All records with disagreement among the 3 reviewers were automatically included in the full-text screening stage. The full-text screening stage was completed by the first author with the help of a research assistant and verified by the fourth author. All disagreements were resolved through discussion.

Stage 4: Charting the Data

For each retrieved record, 2 authors (MEB and MB) extracted the following information into a Microsoft Excel spreadsheet: first author name, year of publication, article title, journal, and number of trial registry (if available), principal investigator name (if available), country of the first author or of the principal investigator (if available). This information was used to identify and map articles pertaining to the same study. In the full-text stage, we also extracted text to verify whether the record included a digital component, was based on a randomized controlled design, focused on specific health conditions, and measured and reported results related to one of the domains of the eHealth literacy model (health, computer, basic or functional, information, media, and scientific literacy).

When multiple records were available for one study, we chose the country of origin of the first author or of the principal investigator listed in the study protocol; we chose the year of publication of the first published article available.

On the basis of the information extracted, we categorized studies according to the domains of the eHealth literacy model (ie, health, computer, basic or functional, information, media, and scientific literacy). We also inductively categorized the studies according to the health conditions described. When multiple conditions were reported, we categorized the study as having multiple conditions. Finally, we inductively categorized the interventions according to four main types of technology: (1) *mobile-based*, including mobile apps, text messages, and interactive voice response, exclusively designed for mobile or other handheld devices; (2) *web-based*, including those designed for being accessed via computer, explicitly labeled as *web- or internet-based*, *online*, and *e-learning*, delivered through bespoke websites or social media outlets, such as social networking sites (eg, Facebook or Twitter)—social media are web-based apps that can be accessed via different devices connected to the internet, including smartphones [39,44,45]; (3) *telehealth*, comprising telerehabilitation, telemedicine, or other interventions focused on distributing services and information via electronic information and telecommunication devices [46]; (4) *EHRs*, focusing on EHRs that are defined as "a repository of patient data in digital form, stored and exchanged securely, and accessible by multiple authorized users" [47]. Telehealth and EHR interventions use the internet to connect various devices, including tablets and mobile phones, yet they represent a different type of delivery mode and format: EHRs. We labeled interventions using a combination of the modes described earlier, as reported in other studies [6,48]. When studies reported a combination of the abovementioned categories, we categorized the study as a *hybrid*.

The first author and a research assistant independently completed the classifications; in case of inconsistencies or disagreements between the classifications, the fourth author acted as a third reviewer and resolved the disagreements through discussion. All the authors agreed with the final categorization.

Stage 5: Collating, Summarizing, and Reporting the Results

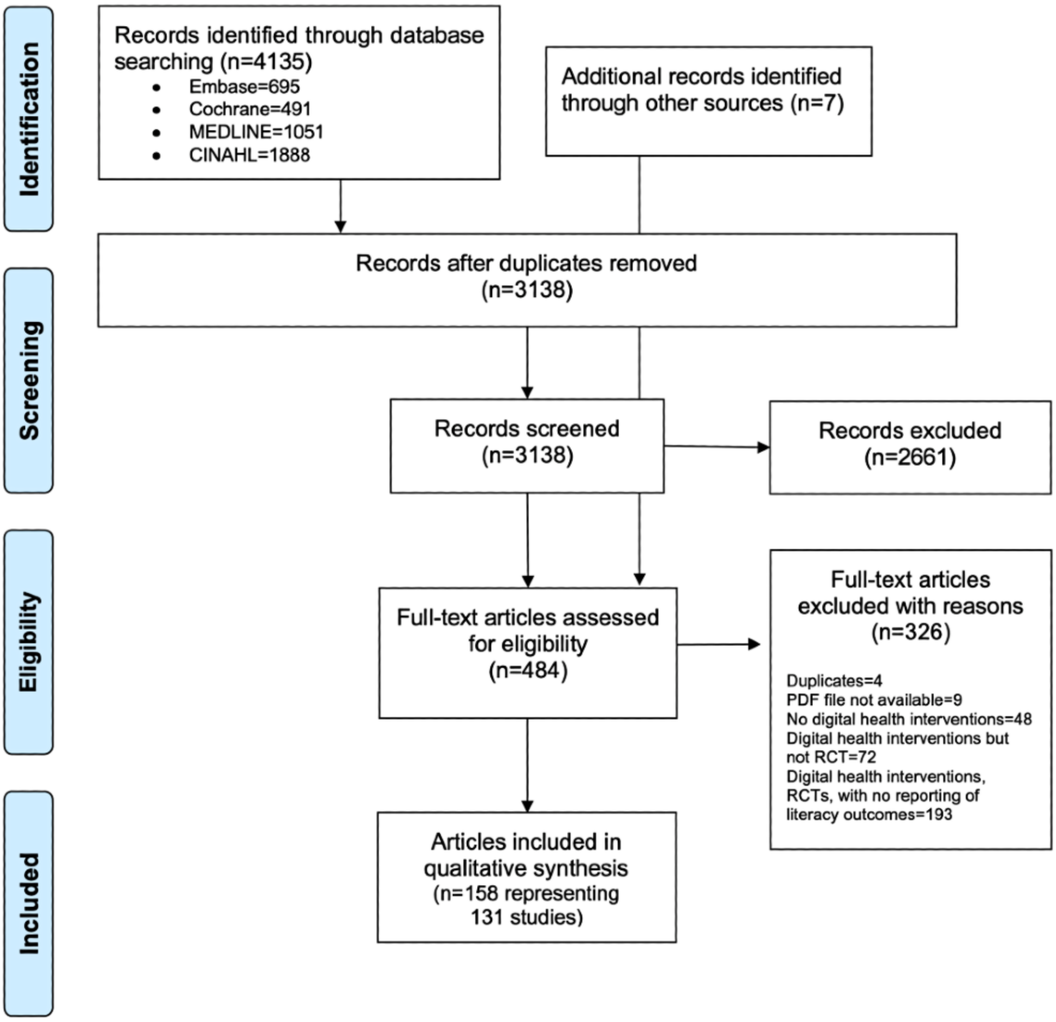
We performed a descriptive analysis of the characteristics of the included papers and reported the results by year of publication, the country of origin of the study authors, eHealth literacy domain, health condition, and type of technology used.

Results

Search Results

The electronic database search yielded 4135 records. The selection process is summarized in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram shown in Figure 1. Briefly, after removing duplicates, the titles and abstracts of 3138 records were independently screened by 3 reviewers. During the title and abstract screening, we excluded 2661 records that were deemed irrelevant. The remaining 477 records were assessed for eligibility in the full text. Scanning the reference lists of two relevant systematic reviews [5,49] allowed us to identify seven other eligible records. We evaluated the eligibility of 484 records that were screened in full text. Of these, 326 records were excluded for the following reasons: 48 did not discuss DHIs (wrong context); 72 reported on digital interventions but did not use an RCT or randomized clinical trial design (wrong study design); 193 records discussed DHIs but did not report any type of literacy (no relevant outcome assessed or reported); 4 were duplicate records; and for the remaining 9 records, we could not retrieve a PDF file. The list of excluded references is provided in Multimedia Appendix 2. Overall, we included 158 records: 79 records reported concluded interventions and 79 records reported protocols of ongoing studies, without reporting results. These were included because they described the assessment of some domains of the eHealth literacy model. The 158 records described a total of 131 unique studies.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. RCT: randomized controlled trial.



Characteristics of the Included Studies

Publication Year and Geographic Distribution

As shown in Figure 2, most of the selected studies were conducted in the last 4 years (86/131, 65.6%), followed by an exponential trend, peaking in 2019 (30/131, 22.9%) and ranging from 2001 to 2020.

The studies were conducted in 26 countries (Table 1). Most studies were conducted across 3 countries (81/131, 61.8%), including the United States (43/131, 32.8%), Australia (28/131, 21.3%), and the United Kingdom (10/131, 7.6%). Approximately one-third of the studies (38/131, 29%) were conducted in European countries such as the United Kingdom (10/38, 26%); Germany (8/38, 21%); Denmark (5/38, 13%); Sweden (4/38, 11%); the Netherlands (3/38, 8%); Norway (2/38, 5%); and Belgium, Finland, Luxemburg, Ireland, Slovakia, and Switzerland (1/38, 3% each). Asian countries were represented by Iran (4/16, 25%); Turkey (3/16, 19%); Hong Kong (2/16, 13%); Singapore (2/16, 13%); Japan (2/16, 13%); and Jordan, Malaysia, and Pakistan (1/16, 6% each). Overall, only 0.8% (1/131) of studies were conducted in Africa (South Africa) and 22.1% (29/131) in Oceania (New Zealand: 1/29, 3%; Australia: 28/29, 97%).

In the following sections, we have reported the results according to our research objectives, whereas a table with the detailed characteristics of the selected studies is provided in Multimedia Appendix 3.

Figure 2. Distribution of selected studies (N=131) by year of publication.

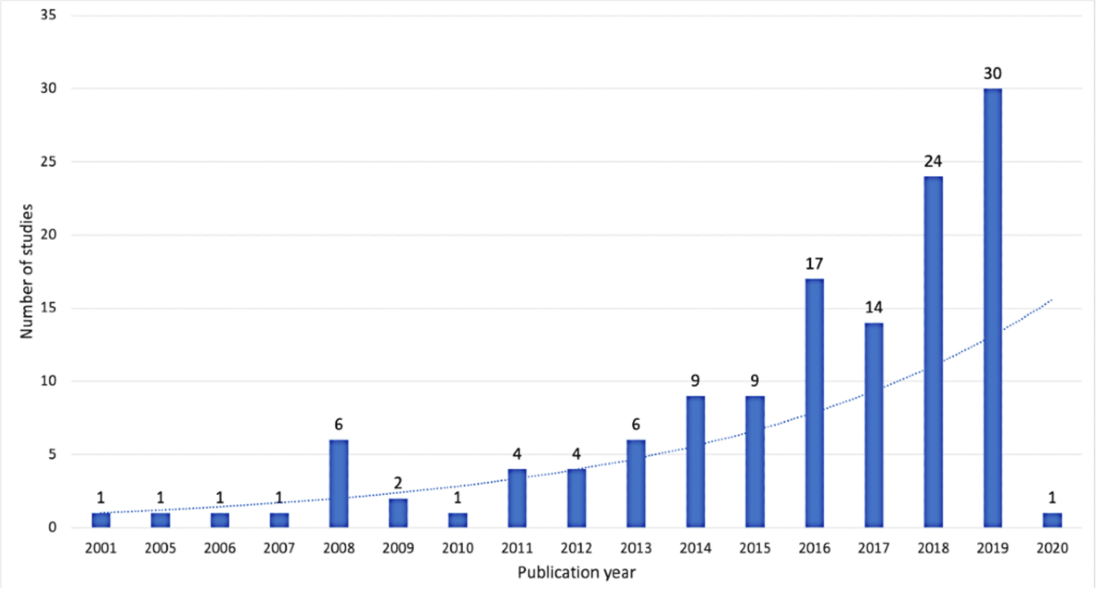


Table 1. Distribution of included studies by country (N=131).

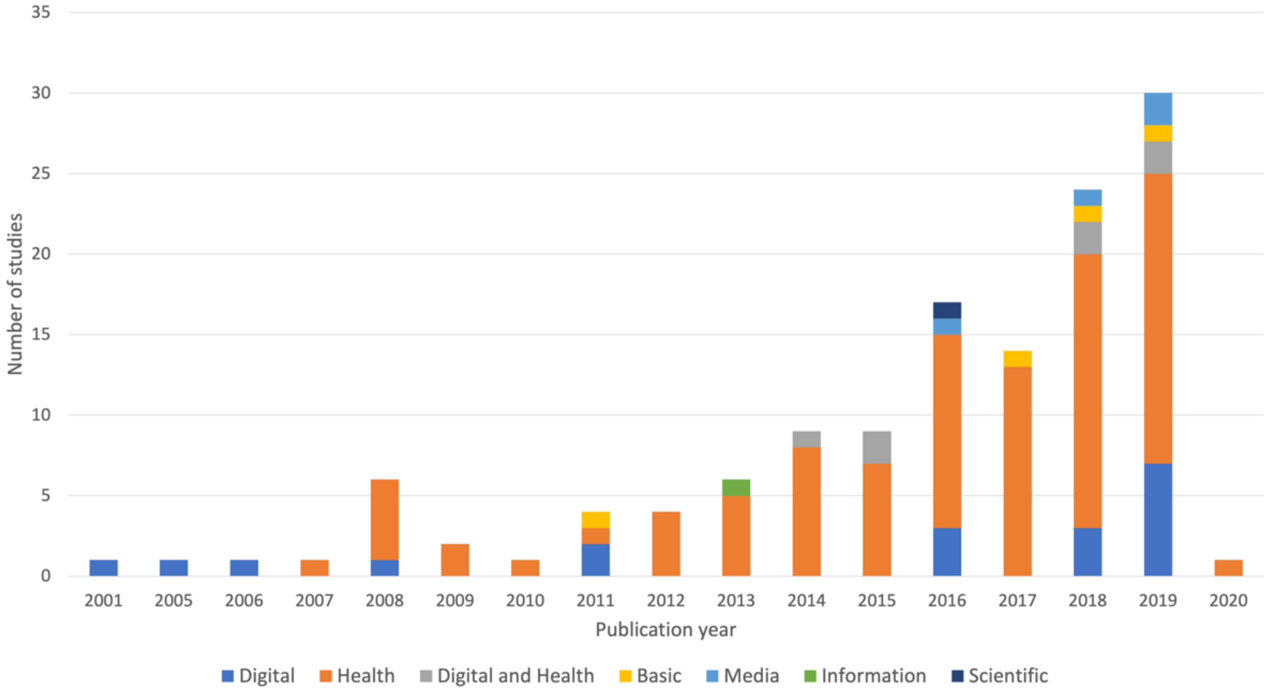
Country	Studies, n (%)
United States	43 (32.8)
Australia	28 (21.3)
United Kingdom	10 (7.6)
Germany	8 (6.1)
Denmark	5 (3.8)
Sweden	4 (3.1)
Iran	4 (3.1)
Netherlands	3 (2.2)
Turkey	3 (2.2)
Brazil	2 (1.5)
Canada	2 (1.5)
Hong Kong	2 (1.5)
Japan	2 (1.5)
Norway	2 (1.5)
Singapore	2 (1.5)
Belgium	1 (0.7)
Finland	1 (0.7)
Ireland	1 (0.7)
Jordan	1 (0.7)
Luxemburg	1 (0.7)
Malaysia	1 (0.7)
New Zealand	1 (0.7)
Pakistan	1 (0.7)
Slovakia	1 (0.7)
South Africa	1 (0.7)
Switzerland	1 (0.7)

Domains of the eHealth Literacy Model Assessed

Figure 3 presents the years of publication of the included studies grouped by domain of the eHealth literacy model. In total, 2.2% (3/131) of the included studies were published before 2006, when the seminal publications of the eHealth literacy model appeared [9,24]. These studies included an assessment of computer literacy. Studies published in 2008 mostly reported assessments of health literacy.

Of the 131 studies included, none assessed or measured all six domains of the eHealth literacy model. Most of the studies (124/131, 94.6%) focused on one of the six domains of the eHealth literacy model; only 5.3% (7/131) of studies reported the assessment of two domains, namely health literacy and digital or computer literacy. Most of the studies that reported on one literacy domain (124/131, 94.7%) focused on health literacy (95/124, 76.6%), followed by digital or computer literacy (19/124, 15.3%), basic or functional literacy (4/124, 3.2%), media literacy (4/124, 3.2%), information literacy (1/124, 0.8%), and scientific literacy (1/124, 0.8%).

Figure 3. Distribution of studies (N=131) by eHealth literacy model domain.



Health Conditions Addressed and Technologies Used

Table 2 provides a summary of the selected studies grouped by technology category and health condition category.

A large number of studies (61/131, 46.5%) discussed interventions addressing NCDs, such as hypertension, obesity, end-stage kidney disease, type 2 diabetes, chronic kidney disease, heart disease (vascular disease, cerebrovascular disorders, ischemic heart disease, coronary artery disease, and heart failure), fibromyalgia syndrome, and asthma. Of these 61 NCD-focused studies, 3 (5%) also discussed mental health topics, and 1 (2%) covered sexual and reproductive health. The second most covered category of health conditions was mental health (26/131, 19.8%), including depression, eating disorders, mental and behavioral disorders, anxiety, and suicide prevention. Other topics included health education (16/131, 12.2%), such as health promotion, health communication, patient provider communication and literacy, aging and maternal and infant health (4/131, 3.0% of studies), sexual and reproductive health, and substance use (3/131, 2.2% of studies each). The remaining 11.4% (15/131) studies covered a variety of health topics.

Table 2. Number of studies by health condition category and type of technology used.

Health condition and technology used	Web based (n=68), n (%)	Mobile based (n=40), n (%)	Telehealth (n=10), n (%)	EHRs[a] (n=5), n (%)	Hybrid (n=8), n (%)	Total (N=131), n (%)
NCDs[b]	22 (32.4)	19 (47.5)	6 (60)	3 (60)	6 (75)	56 (42.7)
NCDs—mental health	1 (1.5)	2 (5)	0 (0)	0 (0)	0 (0)	3 (2.3)
NCDs—sexual and reproductive health	0 (0)	0 (0)	0 (0)	0 (0)	1 (12.5)	1 (0.8)
Mental health	21 (30.9)	4 (10)	1 (10)	0 (0)	0 (0)	26 (19.8)
Aging	2 (2.9)	1 (2.5)	1 (10)	0 (0)	0 (0)	4 (3.1)
Health education topics	9 (13.2)	4 (10)	1 (10)	2 (40)	0 (0)	16 (12.2)
Maternal and infant health	2 (2.9)	2 (5)	0 (0)	0 (0)	0 (0)	4 (3.1)
Sexual and reproductive health	2 (2.9)	1 (2.5)	0 (0)	0 (0)	0 (0)	3 (2.3)
Substance use	2 (2.9)	1 (2.5)	0 (0)	0 (0)	0 (0)	3 (2.3)
Other health topics	7 (10.3)	6 (15)	1 (10)	0 (0)	1 (12.5)	15 (11.5)

[a] EHR: electronic health record.

[b] NCD: noncommunicable disease.

With regard to the technologies used, most studies included web-based interventions (68/131, 51.9%), followed by mobile-based (40/131, 30.5%), telehealth (10/131, 7.6%) EHRs (5/131, 3.8%), and hybrid interventions (8/131, 6.1%). Examples of web-based technology included e-learning portals for specialized training [50,51], experimental websites, and social media platforms [52-54], which are used to deliver motivational or informational campaigns. Mobile-based interventions included health apps [55-57], SMS text messaging or WhatsApp [58], games [59,60], and interactive voice response [61,62]. Telehealth interventions included rehabilitation programs [63,64] or remote counseling [65]. Hybrid interventions included combinations of mobile apps and EHRs [55-57], SMS text messaging and EHRs [66], or a mix of web- and mobile-based technologies [67].

Among web-based interventions (n=68), most focused on NCDs (22/68, 32%), mental health (21/68, 31%), and health education topics (9/68, 13%). Mobile-based interventions (n=40) followed a similar pattern, with approximately half of the studies focusing on NCDs (19/40, 48%) or other health topics (6/40, 15%). Most telehealth (6/10, 60%), EHR (3/5, 60%), and hybrid (6/8, 75%) interventions focused on NCDs.

Discussion

Principal Findings

This is the first scoping review examining the extent to which DHIs have assessed, accounted for, and reported any of the six domains of the eHealth literacy model by Norman and Skinner [9]. We identified a sizable literature discussing DHIs developed in 26 countries, spanning two decades. The eHealth literacy model [9] and eHEALS [24] date back to 2006, but we included 3 studies that were published before that year and all assessed computer literacy. This might indicate that attention toward the ability to use technology was a research interest in the early 2000s. However, this interest has not grown exponentially and concomitantly with the growth of DHIs. It is interesting to observe that the assessment of digital literacy has grown only after 2015, but it has remained below the assessment of health literacy, which was the domain assessed the most over time. There is no clear explanation for these trends. A bibliometric analysis of the studies cited in the seminal papers mentioned above could reveal the connections between publications and demonstrate when the eHealth literacy model has received more citations.

Most of the evidence comes from the *Global North*, that is, from English-speaking countries including the United States, Australia, and the United Kingdom. A few studies have been conducted in countries of the *Global South*, such as Africa, Latin America, or South East Asia. This finding is consistent with that reported in a recent scoping review on digital health innovations [68] and in a recent bibliometric analysis of research on mHealth apps [69], which showed a predominance of articles published in the United States, the United Kingdom, Australia, and Canada. Publication bias and limited evidence from developing countries or the *Global South* has been previously reported in the literature [70-72], yet there seems to be a lack of evidence on DHIs from Africa, the Middle East, South America, or Southeast Asia. There may be various reasons for this absence of evidence. First, research on digital technologies might not have reached an advanced stage to produce interventions with the highest level of evidence (ie, RCTs). Second, the existing digital divide might persist in many countries, both low- and high-income countries [30]; however, mobile phones and telemedicine are becoming more widely adopted [46,73]. Third, researchers based in low- and middle-income countries (LMICs) may be published in languages other than English or might have limited English language proficiency, but this latter assumption does not seem to be grounded in evidence [74,75]. Another reason might be that researchers in LMICs might choose to publish in journals that are not indexed in the databases we searched. Alternatively, researchers in LMICs might not have the possibility to publish their results because of a lack of funding for open access publications or because editors demonstrate publication bias [72]. Regardless of the reasons, we call for digital health researchers based in countries of the Global South to publish more study protocols and diffuse intervention results; we also call the international community of editors and publishing houses to incentivize or support research published from these underrepresented countries, so that stronger conclusions can be drawn from a truly global evidence base.

Domains of the eHealth Literacy Model Assessed

Our findings showed that none of the 131 selected DHIs conducted in the last 20 years accounted for or assessed all six domains of the eHealth literacy model. Although these interventions were included because they assessed at least one domain of the model, only 5.3% (7/131) of studies included the assessment of more than one domain. These 7 studies assessed only digital and health literacy. Our study also shows that most DHIs have assessed and evaluated health literacy [19] among intervention participants, which is an important factor that can determine the health outcomes of a study [21,31,32]. Although the focus on health literacy in DHIs is consistent with some literature reviews combining the study of health literacy identified through our searches [5,22,26,34,35,49], it is somewhat surprising that none of the other four domains of the eHealth literacy model were concomitantly addressed.

There are numerous explanations for these findings. First, researchers specialized in DHIs might not be familiar with or might have ignored the original model, even though the seminal paper by Norman and Skinner [9] and the paper describing the eHEALS [24] are highly cited (as of October 17, 2020, Google Scholar showed 1128 and 1450 citations, respectively). Second, researchers might have decided to focus on other domains of the model while making implicit or explicit assumptions about the levels of literacy in other domains. For example, the limited evidence related to the assessment of the domains of scientific, information, media, and functional literacy might be based on the assumption that digital literacy instruments, such as the popular eHEALS [24], include questions related to the use of information on the internet as a medium of search information; hence, these could be associated with media and information literacy domains. However, there exist several instruments that specifically assess media literacy [76,77], scientific literacy [78,79], and information literacy [80,81]. Moreover, Norman and Skinner [9] did not consider overlapping elements when they developed the eHealth literacy model, which considers the six domains as distinct and separate.

Although intervention designers should aim to develop content that is understood by people with low functional literacy [82,83], this fact should be proven or verified by the same intervention designers. One way to do so is to assess functional literacy or to report the level of literacy rather than to just develop the content of the intervention through formal readability and usability testing. The fact that other domains of the eHealth literacy model were not always conducted raises concerns about the generalizability of such interventions across the eHealth literacy spectrum. DHIs tend to attract tech-savvy, healthy volunteers who have access to technology and who might have different sociodemographic and psychological profiles compared with people who belong to vulnerable segments of the population and do not have access to technology [30].

Another important finding was that few identified DHIs assessed digital literacy (n=26). Not assessing digital literacy is based on the assumption that all participants are equally able to use technology and are able to make sense of the information delivered. This assumption might not be tenable in all contexts, and it does not allow researchers to understand whether participants appropriately received the intervention. In other words, health literacy is context-specific and varies according to different situations and topics. Arguably, health and digital literacy might act as moderators of intervention effects and not including these factors might underestimate or overestimate intervention effects [84].

The limited assessment of digital literacy in DHIs also raises some ethical considerations in terms of equity and social justice, as these interventions tend to attract highly educated, healthy, and digitally literate individuals who have easy access to technology, leaving out less-educated and poorer segments of the population, who may be most in need of the interventions themselves [30,85]. This selection bias isolates segments of the population that are traditionally difficult to reach [86,87], yet it is important to acknowledge that the results of DHIs might be less generalizable than interventions that do not use technology.

Another reason for the absence of a comprehensive and accurate assessment of the six domains of the eHealth literacy model might be due to the fact that this assessment will be unfeasible and daunting for the participants. Holding constant the basic or functional literacy (ie, numeracy and ability to read), assessing all six domains using existing scales for media, scientific, health, digital, and information literacy would require longer questionnaires that will take more time to complete, which might discourage participation in these studies. For example, one of the most used instruments to assess digital literacy is the relatively short (8 items) eHEALS [24]. However, for context-specific domains such as health literacy [9], there are many more instruments available, which vary in length and complexity [25,88]. A recent review identified 43 different instruments [89], and the Health Literacy Toolshed database included 200 measures [90]. Similar issues of measurement pertain to the assessment of literacy in a digital world [91], including media literacy [92]. Nevertheless, we urge digital health researchers to find ways to assess and evaluate the different domains of the eHealth literacy model, so that they can gain a better understanding of the study participants' characteristics, abilities, and needs. If measuring all domains might appear unfeasible, we suggest that DHI researchers prioritize the assessment of digital literacy—using the short eHEALS [24]—and health literacy, which is context specific, according to the model by Norman and Skinner [9]. Once the health topic or context is defined (mental health, breast cancer, etc), the choice of a short, yet valid instrument to assess health literacy in that context would become easier. As digital health and health literacy can change due to the intervention itself, we recommend assessing these constructs before and after the intervention. Finally, media, scientific, and traditional literacy are analytical skills that are not specific to any context; it would be easier for researchers to routinely assess these domains before the start of any intervention.

Health Conditions Addressed and Technologies Used

This scoping review showed that the selected DHIs published in the last 20 years focused mostly on NCDs, delivered via web- or mobile-based platforms. This is consistent with the findings of a few recent scoping reviews focusing on research on DHIs for behavior change [93,94] or in a recent bibliometric analysis of mHealth apps [69]. Although most DHIs have covered NCDs and mental health, there are many avenues for digital health. Further systematic reviews could be developed to specifically qualify and quantify the effectiveness of DHIs delivered via web or mobile phones in reducing NCDs and mental health issues. These systematic reviews could also anticipate sensitivity analyses based on the modes of delivery, length of the interventions, or the assessment of eHealth literacy model domains. This scoping review provides a valuable map of the evidence and sets the research agenda for DHIs in the coming years.

Strengths and Limitations

To the best of our knowledge, this is the first scoping review that systematically examined evidence pertaining to the application of the eHealth literacy model by Norman and Skinner [9] in DHIs. We looked at the highest quality of evidence, following a predetermined search strategy and a systematic approach to appraise the literature, without restricting our searches to specific periods, populations, countries, or health conditions. Nevertheless, this study has some limitations that are common to many other systematic or scoping reviews. These limitations include the fact that we looked only at peer-reviewed articles available in English. It is possible that some evidence on the use of the eHealth literacy model could have been reported in non–peer-reviewed or gray literature. Another limitation is related to the use of an RCT filter and focus on the RCT study design. Although RCTs provide the highest level of evidence, according to Grading of Recommendations Assessment, Development and Evaluation standards [40], it might be possible that some relevant research entailed the use of other types of study designs.

Conclusions

This review suggests that future DHIs should focus more on the assessment of the eHealth literacy domains when developing a DHI, especially the domains that are assessed the least, such as scientific, media, basic, and information literacy. Even though assessing all domains of the eHealth literacy model might be unfeasible, it would allow researchers to account for factors that might moderate or mediate the effects of the interventions on the targeted health outcomes.

Future systematic reviews should be conducted to examine the effects of DHIs on various health outcomes identified in this review by anticipating subgroup or sensitivity analyses comparing different types of intervention, delivery modes, and most importantly different levels of health literacy or digital literacy.

Acknowledgments

The authors wish to thank Mrs Layal Hneiny, a medical librarian at the American University of Beirut, for her help during the development of the search strategies and Ms Fatma Barakat and Ms Dalia Sarieldine for their help with the screening and data selection of included articles.

Authors' Contributions

This study is the result of the Integrative Learning Experience carried out by MEB as a partial fulfillment of the Master of Public Health, offered by the Faculty of Health Sciences, American University of Beirut, Lebanon. TKK and FEJ acted as advisors and second readers of the project, respectively; MB acted as the project’s preceptor and supervisor. MB and MEB conceived and designed the review and MB coordinated it. TKK and FEJ provided intellectual feedback in the development of the study. MEB and MB were involved in developing the search strategy, extracted the data, completed the analyses, and interpreted the data. MEB and MB drafted the manuscript. All authors have reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE search strategy.
[DOCX File , 25 KB - jmir_v23i6e23473_app1.docx]

Multimedia Appendix 2
Records excluded during the full-text screening phase with full citations.
[DOCX File , 74 KB - jmir_v23i6e23473_app2.docx]

Multimedia Appendix 3
Characteristics of included studies with references.
[DOCX File , 265 KB - jmir_v23i6e23473_app3.docx]

References

1. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
2. Soobiah C, Cooper M, Kishimoto V, Bhatia RS, Scott T, Maloney S, et al. Identifying optimal frameworks to implement or evaluate digital health interventions: a scoping review protocol. BMJ Open 2020 Aug 13;10(8):e037643 [FREE Full text] [doi: 10.1136/bmjopen-2020-037643] [Medline: 32792444]
3. Classification of digital health interventions v1.0 : a shared language to describe the uses of digital technology for health. World Health Organization. 2018. URL: https://tinyurl.com/ybd4kedp [accessed 2021-05-17]
4. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, van Roijen LH, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
5. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health Informatics J 2016 Jun;22(2):81-98 [FREE Full text] [doi: 10.1177/1460458214534092] [Medline: 24916567]
6. Bardus M, Smith JR, Samaha L, Abraham C. Mobile phone and web 2.0 technologies for weight management: a systematic scoping review. J Med Internet Res 2015 Nov 16;17(11):e259 [FREE Full text] [doi: 10.2196/jmir.5129] [Medline: 26573984]
7. Car J, Lang B, Colledge A, Ung C, Majeed A. Interventions for enhancing consumers' online health literacy. Cochrane Database Syst Rev 2011 Jun 15(6):CD007092 [FREE Full text] [doi: 10.1002/14651858.CD007092.pub2] [Medline: 21678364]
8. 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]
9. 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]
10. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010;15 Suppl 2:9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
11. 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]
12. Lorini C, Santomauro F, Donzellini M, Capecchi L, Bechini A, Boccalini S, et al. Health literacy and vaccination: a systematic review. Hum Vaccin Immunother 2018 Feb 01;14(2):478-488 [FREE Full text] [doi: 10.1080/21645515.2017.1392423] [Medline: 29048987]
13. Cajita MI, Cajita TR, Han H. Health literacy and heart failure: a systematic review. J Cardiovasc Nurs 2016;31(2):121-130 [FREE Full text] [doi: 10.1097/JCN.0000000000000229] [Medline: 25569150]
14. Taylor DM, Fraser S, Dudley C, Oniscu GC, Tomson C, Ravanan R, ATTOM investigators. Health literacy and patient outcomes in chronic kidney disease: a systematic review. Nephrol Dial Transplant 2018 Sep 01;33(9):1545-1558. [doi: 10.1093/ndt/gfx293] [Medline: 29165627]
15. Ghisi GL, Chaves GS, Britto RR, Oh P. Health literacy and coronary artery disease: a systematic review. Patient Educ Couns 2018 Feb;101(2):177-184. [doi: 10.1016/j.pec.2017.09.002] [Medline: 28899710]
16. Firmino RT, Ferreira FM, Paiva SM, Granville-Garcia AF, Fraiz FC, Martins CC. Oral health literacy and associated oral conditions: a systematic review. J Am Dent Assoc 2017 Aug;148(8):604-613. [doi: 10.1016/j.adaj.2017.04.012] [Medline: 28477838]
17. Zheng M, Jin H, Shi N, Duan C, Wang D, Yu X, et al. The relationship between health literacy and quality of life: a systematic review and meta-analysis. Health Qual Life Outcomes 2018 Oct 16;16(1):201 [FREE Full text] [doi: 10.1186/s12955-018-1031-7] [Medline: 30326903]
18. Michou M, Panagiotakos DB, Costarelli V. Low health literacy and excess body weight: a systematic review. Cent Eur J Public Health 2018 Sep;26(3):234-241 [FREE Full text] [doi: 10.21101/cejph.a5172] [Medline: 30419628]
19. Chinn D. Critical health literacy: a review and critical analysis. Soc Sci Med 2011 Jul;73(1):60-67. [doi: 10.1016/j.socscimed.2011.04.004] [Medline: 21640456]
20. de Wit L, Fenenga C, Giammarchi C, di Furia L, Hutter I, de Winter A, et al. Community-based initiatives improving critical health literacy: a systematic review and meta-synthesis of qualitative evidence. BMC Public Health 2017 Jul 20;18(1):40 [FREE Full text] [doi: 10.1186/s12889-017-4570-7] [Medline: 28728547]
21. Fleary SA, Paasche-Orlow MK, Joseph P, Freund KM. The relationship between health literacy, cancer prevention beliefs, and cancer prevention behaviors. J Cancer Educ 2019 Oct;34(5):958-965 [FREE Full text] [doi: 10.1007/s13187-018-1400-2] [Medline: 30022378]
22. Chesser AK, Woods NK, Smothers K, Rogers N. Health literacy and older adults: a systematic review. Gerontol Geriatr Med 2016;2:A [FREE Full text] [doi: 10.1177/2333721416630492] [Medline: 28138488]
23. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L, et al. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manage E-Learn Int J 2015;7(4):540. [doi: 10.34105/j.kmel.2015.07.035]

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. Collins SA, Currie LM, Bakken S, Vawdrey DK, Stone PW. Health literacy screening instruments for eHealth applications: a systematic review. J Biomed Inform 2012 Jun;45(3):598-607 [FREE Full text] [doi: 10.1016/j.jbi.2012.04.001] [Medline: 22521719]
26. 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]
27. Dominick G, Friedman D, Hoffman-Goetz L. Do we need to understand the technology to get to the science? A systematic review of the concept of computer literacy in preventive health programs. Health Edu J 2010 Mar 15;68(4):296-313. [doi: 10.1177/0017896909349289]
28. Hur I, Lee R, Schmidt J. How healthcare technology shapes health literacy? A systematic review. Healthcare Information Systems and Technology (Sighealth). 2015. URL: https://aisel.aisnet.org/amcis2015/HealthIS/GeneralPresentations/3/ [accessed 2021-05-17]
29. Hæsum LK, Ehlers L, Hejlesen O. Influence of health literacy on outcomes using telehomecare technology: a systematic review. Health Edu J 2015 Jan 09;75(1):72-83. [doi: 10.1177/0017896914565489]
30. Azzopardi-Muscat N, Sørensen K. Towards an equitable digital public health era: promoting equity through a health literacy perspective. Eur J Public Health 2019 Oct 01;29(Supplement_3):13-17 [FREE Full text] [doi: 10.1093/eurpub/ckz166] [Medline: 31738443]
31. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31 Suppl 1:19-26. [doi: 10.5555/ajhb.2007.31.supp.S19] [Medline: 17931132]
32. Neter E, Brainin E. Association between health literacy, ehealth literacy, and health outcomes among patients with long-term conditions. Eur Psychol 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
33. Han H, Hong H, Starbird LE, Ge S, Ford AD, Renda S, et al. Ehealth literacy in people living with HIV: systematic review. JMIR Public Health Surveill 2018 Sep 10;4(3):e64 [FREE Full text] [doi: 10.2196/publichealth.9687] [Medline: 30201600]
34. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
35. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014 Nov 10;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
36. Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth literacy among college students: a systematic review with implications for eHealth education. J Med Internet Res 2011 Dec 01;13(4):e102 [FREE Full text] [doi: 10.2196/jmir.1703] [Medline: 22155629]
37. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
38. Bardus M, Smith JR, Samaha L, Abraham C. Mobile and Web 2.0 interventions for weight management: an overview of review evidence and its methodological quality. Eur J Public Health 2016 Aug;26(4):602-610 [FREE Full text] [doi: 10.1093/eurpub/ckw090] [Medline: 27335330]
39. Bardus M, Rassi RE, Chahrour M, Akl EW, Raslan AS, Meho LI, et al. The use of social media to increase the impact of health research: systematic review. J Med Internet Res 2020 Jul 06;22(7):e15607 [FREE Full text] [doi: 10.2196/15607] [Medline: 32628113]
40. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg 2011 Jul;128(1):305-310 [FREE Full text] [doi: 10.1097/PRS.0b013e318219c171] [Medline: 21701348]
41. Aromataris E, Munn Z, editors. JBI manual for evidence synthesis. In: Joanna Briggs Institute Reviewer's Manual. Australia: JBI - The University of Adelaide; 2020.
42. Munn Z, Peters MD, 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]
43. 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]
44. Constantinides E, Fountain SJ. Web 2.0: Conceptual foundations and marketing issues. J Direct Data Digit Mark Pract 2008 Jan 4;9(3):231-244. [doi: 10.1057/palgrave.dddmp.4350098]
45. Bardus M. The web 2.0 and social media technologies for pervasive health communication: are they effec. Stud Commun Sci 2011;11(1):119-136. [doi: 10.5169/seals-791189]
46. Zobair K, Sanzogni L, Sandhu K, Islam M. Telemedicine adoption opportunities and challenges in the developing world. Eval Chall Opport Healthcare Ref 2020:193. [doi: 10.4018/978-1-7998-2949-2.ch008]
47. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008 May;77(5):291-304. [doi: 10.1016/j.ijmedinf.2007.09.001] [Medline: 17951106]

48. Brørs G, Pettersen TR, Hansen TB, Fridlund B, Hølvold LB, Lund H, et al. Modes of e-Health delivery in secondary prevention programmes for patients with coronary artery disease: a systematic review. BMC Health Serv Res 2019 Jun 10;19(1):364 [FREE Full text] [doi: 10.1186/s12913-019-4106-1] [Medline: 31182100]
49. Carter MC, Burley VJ, Cade JE. Handheld electronic technology for weight loss in overweight/obese adults. Curr Obes Rep 2014 Sep;3(3):307-315. [doi: 10.1007/s13679-014-0112-0] [Medline: 26626760]
50. Taylor-Rodgers E, Batterham PJ. Evaluation of an online psychoeducation intervention to promote mental health help seeking attitudes and intentions among young adults: randomised controlled trial. J Affect Disord 2014 Oct 15;168:65-71. [doi: 10.1016/j.jad.2014.06.047] [Medline: 25038293]
51. Jorm AF, Kitchener BA, Fischer J, Cvetkovski S. Mental health first aid training by e-learning: a randomized controlled trial. Aust N Z J Psychiatry 2010 Dec;44(12):1072-1081. [doi: 10.3109/00048674.2010.516426] [Medline: 21070103]
52. Tse CK, Bridges SM, Srinivasan DP, Cheng BS. Social media in adolescent health literacy education: a pilot study. JMIR Res Protoc 2015 Mar 09;4(1):e18 [FREE Full text] [doi: 10.2196/resprot.3285] [Medline: 25757670]
53. Feinkohl I, Flemming D, Cress U, Kimmerle J. The impact of personality factors and preceding user comments on the processing of research findings on deep brain stimulation: a randomized controlled experiment in a simulated online forum. J Med Internet Res 2016 Mar 03;18(3):e59 [FREE Full text] [doi: 10.2196/jmir.4382] [Medline: 26940848]
54. Hui A, Wong PW, Fu K. Evaluation of an online campaign for promoting help-seeking attitudes for depression using a facebook advertisement: an online randomized controlled experiment. JMIR Ment Health 2015;2(1):e5 [FREE Full text] [doi: 10.2196/mental.3649] [Medline: 26543911]
55. Schougaard LM, Mejdahl CT, Petersen KH, Jessen A, de Thurah A, Sidenius P, et al. Effect of patient-initiated versus fixed-interval telePRO-based outpatient follow-up: study protocol for a pragmatic randomised controlled study. BMC Health Serv Res 2017 Jan 26;17(1):83 [FREE Full text] [doi: 10.1186/s12913-017-2015-8] [Medline: 28122609]
56. Navaneethan SD, Jolly SE, Schold JD, Arrigain S, Nakhoul G, Konig V, et al. Pragmatic randomized, controlled trial of patient navigators and enhanced personal health records in CKD. Clin J Am Soc Nephrol 2017 Sep 07;12(9):1418-1427 [FREE Full text] [doi: 10.2215/CJN.02100217] [Medline: 28778854]
57. Redfern J, Usherwood T, Harris MF, Rodgers A, Hayman N, Panaretto K, et al. A randomised controlled trial of a consumer-focused e-health strategy for cardiovascular risk management in primary care: the Consumer Navigation of Electronic Cardiovascular Tools (CONNECT) study protocol. BMJ Open 2014 Jan 31;4(2):e004523 [FREE Full text] [doi: 10.1136/bmjopen-2013-004523] [Medline: 24486732]
58. Khaleghi M, Shokravi FA, Peyman N, Moridi M. Evaluating the effect of educational interventions on health literacy through social networking services to promote students' quality of life. Korean J Fam Med 2019 May;40(3):188-193 [FREE Full text] [doi: 10.4082/kjfm.18.0060] [Medline: 30997783]
59. Parisod H, Pakarinen A, Axelin A, Löyttyniemi E, Smed J, Salanterä S. Feasibility of mobile health game "Fume" in supporting tobacco-related health literacy among early adolescents: a three-armed cluster randomized design. Int J Med Inform 2018 May;113:26-37. [doi: 10.1016/j.ijmedinf.2018.02.013] [Medline: 29602430]
60. Sanchez R, Brown E, Kocher K, DeRosier M. Improving children's mental health with a digital social skills development game: a randomized controlled efficacy trial of adventures aboard the S.S. GRIN. Games Health J 2017 Feb;6(1):19-27. [doi: 10.1089/g4h.2015.0108] [Medline: 28051877]
61. Cyan R, Hussain H, Shehki I, Ishaq Z, Khan T, Rothenberg R. Affordable technology for saving maternal and infant lives: moving on with solutions. Ann Glob Health 2016;82(3):382. [doi: 10.1016/j.aogh.2016.04.620]
62. Kamal AK, Muqeet A, Farhat K, Khalid W, Jamil A, Gowani A, et al. Using a tailored health information technology-driven intervention to improve health literacy and medication adherence in a Pakistani population with vascular disease (Talking Rx) - study protocol for a randomized controlled trial. Trials 2016 Mar 05;17(1):121 [FREE Full text] [doi: 10.1186/s13063-016-1244-1] [Medline: 26944938]
63. Rezvani F, Härter M, Dirmaier J. Promoting a home-based walking exercise using telephone-based health coaching and activity monitoring for patients with intermittent claudication (TeGeCoach): protocol for a randomized controlled trial. Psychother Psych Med 2018 Aug 06;68(08):e43. [doi: 10.1055/s-0038-1667988]
64. Dinesen B, Dittmann L, Gade JD, Jørgensen CK, Hollingdal M, Leth S, et al. "Future Patient" telerehabilitation for patients with heart failure: protocol for a randomized controlled trial. JMIR Res Protoc 2019 Sep 19;8(9):e14517 [FREE Full text] [doi: 10.2196/14517] [Medline: 31538944]
65. Mudiyanselage SB, Stevens J, Watts JJ, Toscano J, Kotowicz MA, Steinfort CL, et al. Personalised telehealth intervention for chronic disease management: a pilot randomised controlled trial. J Telemed Telecare 2019 Jul;25(6):343-352. [doi: 10.1177/1357633X18775850] [Medline: 29793387]
66. Sexual Health Empowerment for Women's Health. Clinical Trials. 2019. URL: https://clinicaltrials.gov/ct2/show/NCT03984695 [accessed 2021-05-17]
67. Duncan M, Vandelanotte C, Kolt GS, Rosenkranz RR, Caperchione CM, George ES, et al. Effectiveness of a web- and mobile phone-based intervention to promote physical activity and healthy eating in middle-aged males: randomized controlled trial of the ManUp study. J Med Internet Res 2014 Jun 12;16(6):e136 [FREE Full text] [doi: 10.2196/jmir.3107] [Medline: 24927299]

68. Iyawa G, Herselman M, Botha A. A scoping review of digital health innovation ecosystems in developed and developing countries. In: Proceedings of the IST-Africa Week Conference (IST-Africa). 2017 Presented at: IST-Africa Week Conference (IST-Africa); May 30- June 2, 2017; Windhoek, Namibia. [doi: 10.23919/istafrica.2017.8102325]
69. Peng C, He M, Cutrona SL, Kiefe CI, Liu F, Wang Z. Theme trends and knowledge structure on mobile health apps: bibliometric analysis. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e18212 [FREE Full text] [doi: 10.2196/18212] [Medline: 32716312]
70. Singh D. Publication bias - a reason for the decreased research output in developing countries. Afr J Psych 2006 Oct 02;9(3):153-155. [doi: 10.4314/ajpsy.v9i3.30216]
71. Mulimani P. Publication bias towards Western populations harms humanity. Nat Hum Behav 2019 Oct 10;3(10):1026-1027. [doi: 10.1038/s41562-019-0720-5] [Medline: 31602030]
72. Yousefi-Nooraie R, Shakiba B, Mortaz-Hejri S. Country development and manuscript selection bias: a review of published studies. BMC Med Res Methodol 2006 Aug 01;6:37 [FREE Full text] [doi: 10.1186/1471-2288-6-37] [Medline: 16879753]
73. Adelakun O, Garcia R. Technical factors in telemedicine adoption in extreme resource-poor countries. In: Global Health and Volunteering Beyond Borders. Switzerland: Springer; 2019:83-101.
74. Hyland K. Academic publishing and the myth of linguistic injustice. J Second Lang Writ 2016 Mar;31:58-69. [doi: 10.1016/j.jslw.2016.01.005]
75. Hyland K. Participation in publishing: the demoralizing discourse of disadvantage. In: Habibie P, Hyland K, editors. Novice Writers and Scholarly Publication. Switzerland: Palgrave Macmillan; 2018:13-33.
76. Ashley S, Maksl A, Craft S. Developing a news media literacy scale. Journal Mass Commun Educ 2013 Jan 02;68(1):7-21. [doi: 10.1177/1077695812469802]
77. Eristi B, Erdem C. Development of a media literacy skills scale. Contemp Educ Technol 2017;8(3):A. [doi: 10.30935/cedtech/6199]
78. Bybee R, McCrae B, Laurie R. PISA 2006: an assessment of scientific literacy. J Res Sci Teach 2009 Oct;46(8):865-883. [doi: 10.1002/tea.20333]
79. Harlen W. The assessment of scientific literacy in the OECD/PISA project. Stud Sci Educ 2001 Jan;36(1):79-103. [doi: 10.1080/03057260108560168]
80. Kurbanoglu SS, Akkoyunlu B, Umay A. Developing the information literacy self - efficacy scale. J Doc 2006 Nov;62(6):730-743. [doi: 10.1108/00220410610714949]
81. Çoklar AN, Yaman ND, Yurdakul IK. Information literacy and digital nativity as determinants of online information search strategies. Comp Hum Behav 2017 May;70:1-9. [doi: 10.1016/j.chb.2016.12.050]
82. Mackert M, Kahlor L, Tyler D, Gustafson J. Designing e-health interventions for low-health-literate culturally diverse parents: addressing the obesity epidemic. Telemed J E Health 2009 Sep;15(7):672-677 [FREE Full text] [doi: 10.1089/tmj.2009.0012] [Medline: 19694596]
83. Mackert M, Champlin SE, Holton A, Muñoz II, Damásio MJ. eHealth and health literacy: a research methodology review. J Comput-Mediat Comm 2014 Apr 12;19(3):516-528. [doi: 10.1111/jcc4.12044]
84. Pelikan JM, Ganahl K, Roethlin F. Health literacy as a determinant, mediator and/or moderator of health: empirical models using the European Health Literacy Survey dataset. Glob Health Promot 2018 Nov 14:57-66. [doi: 10.1177/1757975918788300] [Medline: 30427258]
85. Bardus M, Hassan FA. eHealth for obesity prevention among low-income populations: is research is promoting health for all? Marco Bardus. Eur J Public Health 2016;26(1):77. [doi: 10.1093/eurpub/ckw166.010]
86. Glasgow RE. eHealth evaluation and dissemination research. Am J Prev Med 2007 May;32(5 Suppl):119-126. [doi: 10.1016/j.amepre.2007.01.023] [Medline: 17466816]
87. Atienza AA, Hesse BW, Baker TB, Abrams DB, Rimer BK, Croyle RT, et al. Critical issues in eHealth research. Am J Prev Med 2007 May;32(5 Suppl):71-74 [FREE Full text] [doi: 10.1016/j.amepre.2007.02.013] [Medline: 17466821]
88. Haun J, Luther S, Dodd V, Donaldson P. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. J Health Commun 2012;17 Suppl 3:141-159. [doi: 10.1080/10810730.2012.712615] [Medline: 23030567]
89. Duell P, Wright D, Renzaho AM, Bhattacharya D. Optimal health literacy measurement for the clinical setting: a systematic review. Patient Educ Couns 2015 Nov;98(11):1295-1307. [doi: 10.1016/j.pec.2015.04.003] [Medline: 26162954]
90. Health Literacy Tool Shed. URL: https://healthliteracy.bu.edu/ [accessed 2021-05-17]
91. Jones-Kavalier BR, Flannigan SL. Connecting the digital dots : literacy of the 21st century. Educause Q. 2006. URL: https://er.educause.edu/-/media/files/articles/2006/4/eqm0621.pdf?la=en&hash=072FD518E173153E68975F73DA646847D9736291 [accessed 2021-05-17]
92. Ptaszek G. Media literacy outcomes, measurement. In: The International Encyclopedia of Media Literacy. Hoboken, New Jersey, United States: Wiley; 2019.
93. Joseph-Shehu EM, Ncama BP, Mooi N, Mashamba-Thompson TP. The use of information and communication technologies to promote healthy lifestyle behaviour: a systematic scoping review. BMJ Open 2019 Oct 28;9(10):e029872 [FREE Full text] [doi: 10.1136/bmjopen-2019-029872] [Medline: 31662364]

94. Taj F, Klein MC, van Halteren A. Digital health behavior change technology: bibliometric and scoping review of two decades of research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311 [FREE Full text] [doi: 10.2196/13311] [Medline: 31833836]

Abbreviations

DHI: digital health intervention
eHEALS: eHealth literacy scale
EHR: electronic health record
LMIC: low- and middle-income country
mHealth: mobile health
NCD: noncommunicable disease
PCC: Population-Concept-Context
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 13.08.20; peer-reviewed by J Dubin, Z Wang, O El-Gayar; comments to author 30.09.20; revised version received 20.10.20; accepted 14.04.21; published 03.06.21.

Please cite as:
El Benny M, Kabakian-Khasholian T, El-Jardali F, Bardus M
Application of the eHealth Literacy Model in Digital Health Interventions: Scoping Review
J Med Internet Res 2021;23(6):e23473
URL: https://www.jmir.org/2021/6/e23473
doi:10.2196/23473
PMID:34081023

Review

Characteristics of Antivaccine Messages on Social Media: Systematic Review

Dominik Wawrzuta[1], MD; Mariusz Jaworski[1], PhD; Joanna Gotlib[1], PhD; Mariusz Panczyk[1], PhD

[1] Department of Education and Research in Health Sciences, Faculty of Health Sciences, Medical University of Warsaw, Warsaw, Poland

Corresponding Author:
Dominik Wawrzuta, MD
Department of Education and Research in Health Sciences
Faculty of Health Sciences
Medical University of Warsaw
ul Żwirki i Wigury 81
Warsaw, 02-091
Poland
Phone: 48 225720490
Email: dwawrzuta@wum.edu.pl

Abstract

Background: Supporters of the antivaccination movement can easily spread information that is not scientifically proven on social media. Therefore, learning more about their posts and activities is instrumental in effectively reacting and responding to the false information they publish, which is aimed at discouraging people from taking vaccines.

Objective: This study aims to gather, assess, and synthesize evidence related to the current state of knowledge about antivaccine social media users' web-based activities.

Methods: We systematically reviewed English-language papers from 3 databases (Scopus, Web of Science, and PubMed). A data extraction form was established, which included authors, year of publication, specific objectives, study design, comparison, and outcomes of significance. We performed an aggregative narrative synthesis of the included studies.

Results: The search strategy retrieved 731 records in total. After screening for duplicates and eligibility, 18 articles were included in the qualitative synthesis. Although most of the authors analyzed text messages, some of them studied images or videos. In addition, although most of the studies examined vaccines in general, 5 focused specifically on human papillomavirus vaccines, 2 on measles vaccines, and 1 on influenza vaccines. The synthesized studies dealt with the popularity of provaccination and antivaccination content, the style and manner in which messages about vaccines were formulated for the users, a range of topics concerning vaccines (harmful action, limited freedom of choice, and conspiracy theories), and the role and activity of bots in the dissemination of these messages in social media.

Conclusions: Proponents of the antivaccine movement use a limited number of arguments in their messages; therefore, it is possible to prepare publications clarifying doubts and debunking the most common lies. Public health authorities should continuously monitor social media to quickly find new antivaccine arguments and then create information campaigns for both health professionals and other users.

(J Med Internet Res 2021;23(6):e24564) doi:10.2196/24564

KEYWORDS

vaccination; social media; antivaccination movement; vaccination refusal; health communication; public health; vaccines

Introduction

Background

From the time of its inception, the internet has provided supporters of the antivaccine movement with unprecedented possibilities to affect societies by nearly global communication of their opinions and convictions. The internet has immense potential as a medium for disseminating health information. However, this information can be unreferenced, incomplete, or informal, and thus, it can be considered dangerous [1-5]. The antivaccination messages on the internet are far more unbridled than in other media. For this reason, the internet is a source that can lead to the risk of people making uninformed decisions about vaccination [6]. Contemporary social media is regarded as a major communication tool for spreading information about the antivaccination movement [7,8].

Early studies concerning the content shared on antivaccination websites revealed that this thematic area includes 3 main issues: concerns about vaccine safety and effectiveness, concerns about governmental abuses, and a preference for alternative health practices [2,6,9,10]. First, a key aspect is the perceived risk of severe and long-term side effects of vaccinating a child. Second, mandatory vaccinations are perceived as an unacceptable breach of the right to make independent decisions and as a limitation of civil liberties. A mistrust also exists that is expressed through the belief that governmental supervision bodies keep the reports about the adverse reactions associated with vaccinations a secret and collaborate with the pharmaceutical industry to profit from vaccine sales. Third, alternative health practices are promoted as being more natural than conventional medicine, and they are believed to eliminate the need for vaccination once they are properly followed.

The majority of the antivaccine arguments evoke negative emotions such as fear [8]. In contrast, many provaccine arguments and campaigns are grounded in the values of harm and fairness [11]. Furthermore, the moral ideas of purity and liberty are mostly associated with vaccine hesitancy. The people for whom these values are of great significance will be insensitive to the information on the positive aspects of vaccinations [12]. Most of the arguments used by antivaccination activists can be perceived as part of a broader phenomenon called *denialism*. Denialism can be defined as “the employment of rhetorical arguments to give the appearance of a legitimate debate where there is none, an approach that has the ultimate goal of rejecting a proposition on which a scientific consensus exists” [13]. The European Center for Disease Prevention and Control defines vaccine hesitancy as a “delay in acceptance or refusal of vaccines despite availability of vaccination services” [14]. Such a delay results from broader influences and should always be considered in the historical, political, and sociocultural contexts in which vaccinations occur [15-17]. According to the World Health Organization, vaccine hesitancy is a continuum with the people who are unsure about vaccines or want to delay only some vaccinations being on one end and people who refuse all vaccines being on the other [18].

The spread of negative information about vaccination on the internet and social media is considered to be the leading cause of vaccine hesitancy [17]. Many studies have revealed the negative impact of media controversies related to vaccination safety on the level of vaccination coverage [19,20]. The ubiquitous presence of antivaccination content on the World Wide Web contributes to the prompt dissemination of rumors, myths, and false opinions about vaccines, which subsequently lowers vaccination coverage [4,10,21]. The results of a study by Betsch et al [22] revealed that reading antivaccination webpages for even approximately 5-10 minutes negatively affects the perception of the risk related to vaccination. Dube et al [23] and Smith et al [24] made similar observations. Glanz et al [7] emphasized that exposure to antivaccine messages through social media may intensify parents’ worries and change their intention to vaccinate their children.

Web 2.0 functions (such as those in Facebook, Twitter, YouTube, and blogs and discussion groups) enable users to develop and share content; thus, users can easily and quickly share their personal experiences related to vaccination [21,25]. In this context, social media has become a platform where intimate and emotionally loaded information is exchanged [26]. Here, one can also find the support of people with similar health problems or doubts [27]. Some studies have revealed that content related to vaccines can be found in all social media and that the critical content generally correlated with the content previously published on an antivaccination websites [28,29]. Several important claims on social media may have promoted this phenomenon, including claims that social media is a *big business* and is powerful, that the audience is poorly defined, that fairness is irrelevant, and that nothing is truly private [30]. These factors may affect people’s decisions concerning treatment and preventive health care, including preventive vaccination.

Searching for health information on social media has become increasingly popular [26,31]. Analyses revealed that 41% of parents found antivaccination content through social media [32]. Nearly 16% of all internet users look for information about vaccinations, and 70% of them confirmed that the information they had found affected their decisions [33]. Evrony and Caplan [8] emphasized the need for a more effective search and criticism of the false information on vaccination published on the internet. They highlighted that although every internet user has a choice, spreading disinformation and distorting the facts concerning vaccines is not a choice. Such activities should be noted and effectively challenged. Therefore, it is crucial to learn not only about antivaccine information but also about its emotional load and to identify the content addressees. Such a thorough analysis will help to better understand the motivations and emotions that accompany the information created by antivaccinationists.

Currently, the literature is limited to review articles on the potential role of social media in influencing vaccination beliefs and behavior [10,21,33]. Despite the growing number of papers studying the content published by supporters of the antivaccination movement in social media in recent years, no systematic review analyzing these papers has been developed so far. Previous literature reviews have analyzed disinformation concerning health on social media [34] and health information shared on YouTube [35]. They tackled the issue of vaccination and antivaccination movements only on a fragmented basis, as part of a broader topic. Catalan-Matamoros and Peñafiel-Saiz [36] published a systematic review investigating the communication about vaccinations in traditional media, excluding the content shared on the internet and social media. Considering the activity of the antivaccination movement’s supporters on social media and how easily they can communicate their messages that are not scientifically confirmed to a large number of recipients, it is crucial to learn and understand their activities and messages. This knowledge will help us react and respond effectively to the false information they publish.

Objectives

This study aims to gather, assess, and synthesize evidence regarding the current state of knowledge about antivaccine social media users' web-based activities. We checked 4 general hypotheses about the features of antivaccine content on social media that are often described in papers in the field of vaccinology:

1. Antivaccine messages are more liked and shared than provaccine content.
2. Antivaccine content is more user friendly than provaccine content.
3. Antivaccine social media users describe vaccines as harmful for health or ineffective.
4. Antivaccine social media users share conspiracy theories or claims that are not scientifically proven.

In addition, we summarized the current knowledge on the role of social media bots in the spread of antivaccine messages.

Methods

Design

We designed a systematic review that systematically searched, appraised, and synthesized research evidence following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [37].

Eligibility Criteria

We included papers based on the following inclusion criteria:

1. Papers based only on original studies.
2. Papers analyzing the user-generated content available on social media platforms (including Twitter, Facebook, Instagram, Pinterest, YouTube, LinkedIn, TikTok, Snapchat, VKontakte, WhatsApp, WeChat, Tumblr, Qzone, Reddit, and Sina Weibo).
3. Papers in which the authors extracted a group of antivaccine messages and conducted analyses that contained both qualitative and quantitative elements.

The exclusion criteria were as follows:

1. Papers not written in English.
2. Papers in which vaccines were not the main topic of research.
3. Methodological papers.
4. Conference abstracts.

Information Sources and Search Strategies

Our search strategy was guided by the following research question: *What are the existing data about antivaccine users' activity on social media*? We conducted a scoping literature search of papers published between January 1, 2015, and December 31, 2019, on 3 databases (Scopus, Web of Science, and PubMed), taking into account the PRISMA guidelines. Overall, 2 reviewers (DW and MP) independently screened the databases using the same protocol. The search was conducted in January 2020. We chose the last 5 years for two reasons. First, in 2014, the social media market began to resemble the one known today. The growth of the major platform Facebook slowed, but the activity of users increased. Furthermore, smaller platforms such as Twitter, Instagram, Pinterest, and LinkedIn gained popularity [38]. Second, from our abstract analysis, we found no articles published in 2014 that were potentially eligible for this systematic review; this suggests that the number of papers analyzing negative messages on social media started to grow in 2015.

The query used in the search engines of the databases is as follows:

(("social media" OR "Twitter" OR "Facebook" OR "Instagram" OR "Pinterest" OR "YouTube" OR "LinkedIn" OR "TikTok" OR "Snapchat" OR "VKontakte" OR

"WhatsApp" OR "WeChat" OR "Tumblr" OR "Qzone" OR "Reddit" OR "Sina Weibo")

AND

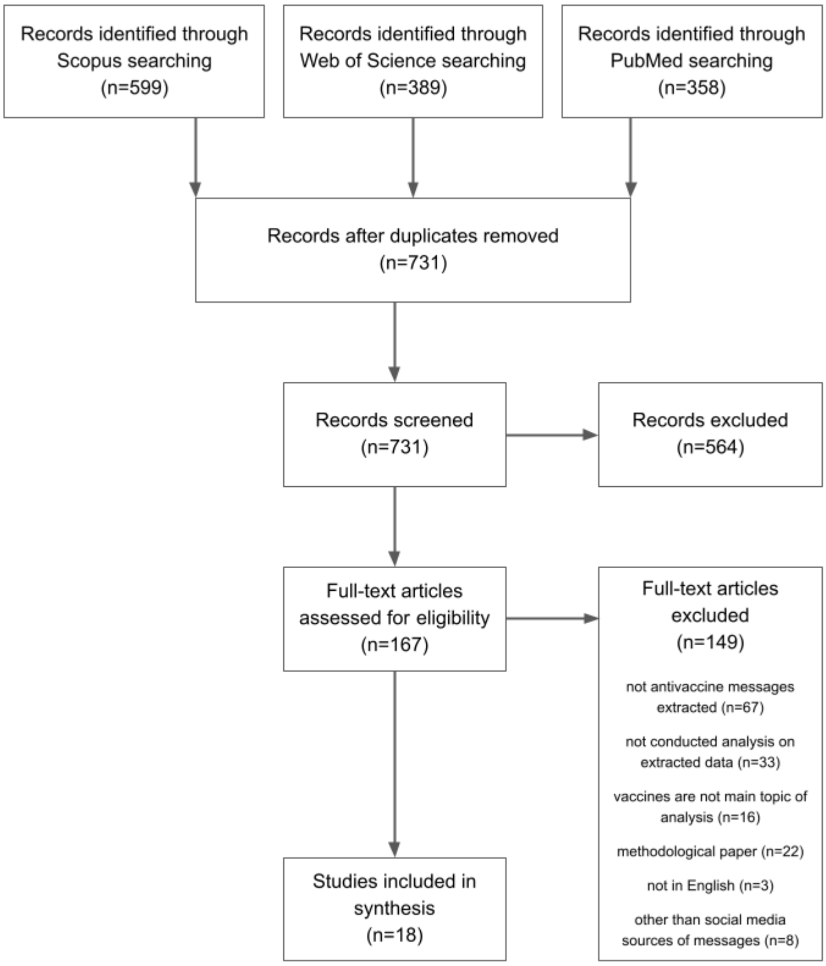
("vaccin*" OR "immuni*"))

We also searched the reference lists of the previously retrieved studies and literature reviews.

Study Selection

A 3-stage approach was used to include and exclude studies in the final review process. Initially, duplicate studies were excluded; subsequently, a screening was performed based on the (1) title, (2) abstract, and finally, (3) full text. Discrepancies regarding article selection were resolved by a consensus within the reviewing team (Figure 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.



Data Collection Process, Extraction, and Quality Assessment

We found 599 articles from Scopus, 389 from Web of Science, and 358 from PubMed. After removing duplicates, 731 articles were analyzed further. In the next step, the titles, abstracts, and keywords were screened. At this stage, we removed papers dealing with pharmacovigilance on social media, survey research, theoretical mathematical models, interventions in social media to increase vaccination coverage, and descriptive analysis without the messages being extracted. Two researchers (DW and MP) conducted a full-text analysis of the remaining 167 papers. A data extraction form was established, and it included authors, year of publication, country, specific objectives, study design, social media platform, type of data analyzed, number of messages studied, main objectives and findings, and geographic focus. Standardized checklist tools were not used to assess the quality and risk of bias of individual studies.

Synthesis of the Results

The outcomes of the 18 studies included in the final analysis were heterogeneous. We performed an aggregative narrative synthesis of the included studies.

Results

Studies' Characteristics

The number of articles analyzing antivaccination messages on social media has increased over the last 5 years. This topic became highly popular in 2019, during which 7 of the 18 included papers were published. Overall, 7 of the 18 papers studied Twitter. YouTube was slightly less popular and was analyzed 6 times. The remaining articles studied Facebook, Instagram, and Pinterest. Although most of the authors analyzed text messages, some of them studied images or videos. The number of analyzed messages varied from 123 to 1,793,690 across the papers. The oldest messages were published in 2006. Table 1 describes the general characteristics of the included studies.

Table 1. Characteristics of the included studies.

Study	Platform	Type of data	Number of messages	Geographic focus	Main topic	Time interval
Basch and MacLean (2019) [39]	Instagram	Text, image	150	English language	HPV[a]	2018
Blankenship et al (2018) [40]	Twitter	Text	1626	English language	Not specified	2014-2015
Broniatowski et al (2018) [41]	Twitter	Text	1,793,690	English language	Not specified	2014-2017
Covolo et al (2017) [42]	YouTube	Video	123	Italian language	Not specified	2014-2015
Donzelli et al (2018) [43]	YouTube	Video	560	Italian language	Not specified	2007-2017
Ekram et al (2019) [44]	YouTube	Text, video	35 videos with comments	English language	HPV	2014
Faasse et al (2016) [45]	Facebook	Text	1489	English language	Not specified	2016
Guidry et al (2015) [46]	Pinterest	Images	800	English language	Not specified	2014
Harvey et al (2019) [47]	Facebook	Meme	234	English language	Not specified	Not specified
Kang et al (2017) [48]	Twitter	Text	26,389	United States	Not specified	2015
Kearney et al (2019) [49]	Instagram	Text, image, video	360	English language	HPV	2011-2018
Luo et al (2019) [50]	Twitter	Text	287,100	English language	HPV	2008-2017
Massey et al (2016) [51]	Twitter	Text	193,379	English language	HPV	2014-2015
Mitra et al (2016) [52]	Twitter	Text	315,240	English language	Not specified	2012-2015
Song and Gruzd (2017) [53]	YouTube	Video	1984	English language	Not specified	Not specified
Venkatraman et al (2015) [54]	YouTube	Video	175	English language	Not specified	Not specified
Yiannakoulias et al (2019) [55]	YouTube	Video	275	English language	Influenza, measles	2006-2018
Yuan et al (2019) [56]	Twitter	Text	669,136	English language	Measles	2015

[a] HPV: human papillomavirus.

There are differences in the methods used to extract the messages. Most of the authors searched for general vaccine-related queries [41,42,46-48,52] or the content of tags [40]. Papers analyzing the human papillomavirus (HPV) vaccine used both general terms (such as *HPV*, *HPV vaccine*, *#HPV*, and *#HPVvaccine*) [39,44,49-51] and specific vaccine names (*Gardasil* and *Cervarix*) [49,51]. Polarized opinions, both positive and negative, were isolated using phrases containing words connected to vaccination and autism [43,53,54]. As Facebook does not allow for searching of posts using a search engine, the authors analyzing this platform used more sophisticated methods to extract messages. Faasse et al [45] studied responses to a photograph that promoted vaccination and was published by Facebook CEO Mark Zuckerberg. Harvey et al [47] found 128 Facebook fan pages devoted to vaccination and sampled up to 30 memes from each page. The papers included content analyzed at different time intervals (Table 1). Only 1 study used search criteria specifically related to the outbreak of an infectious disease. Yuan et al [56] used Twitter data from February 1 to March 9, 2015, to study the tweets published just after the measles outbreak in Disneyland, California, in 2015.

Although most studies examined vaccines in general, 5 focused specifically on HPV vaccines, 2 on measles, and 1 on influenza. Among the articles examining negative comments about the HPV vaccine, 2 examined Twitter, 2 Instagram, and 1 YouTube. These HPV-related messages mention the side effects and risks associated with vaccination [50,51] but often omit some information [44]. Furthermore, HPV-related antivaccine messages on Instagram used more sophisticated forms of communication, including videos or text with images [49]. Provaccine content describing the HPV vaccine contained information on protection against and prevention of cancer [39,50,51] and the safety of the vaccine [44].

In Table 2, we have aggregated the research questions of the included studies into 4 general hypotheses. In the next paragraphs, we describe the results of the analyzed articles in terms of these hypotheses and we discuss bots' activities on social media.

Table 2. Results of the included studies.

Study	Hypothesis 1: Antivaccine messages are more liked and shared than provaccine content	Hypothesis 2: Antivaccine users share more user-friendly content than provaccine users	Hypothesis 3: Antivaccine users describe vaccines as harmful for health	Hypothesis 4: Antivaccine users share conspiracy theories or claims that are not scientifically proven
Basch and MacLean (2019) [39]	Antivaccine messages have more likes than provaccine content (*P*=.02)	N/A[a]	Antivaccine users are less likely to mention the protection provided by the HPV[b] vaccine (*P*<.001)	N/A
Blankenship et al (2018) [40]	Antivaccine messages have more retweets than provaccine ones (*P*<.001)	Most shared antivaccine-related URLs link to Twitter, YouTube, and Facebook	N/A	N/A
Broniatowski et al (2018) [41]	N/A	Content polluters, which may have been primarily created for marketing, spread more antivaccine content than the average nonbot users (*P*<.001)	N/A	N/A
Covolo et al (2017) [42]	Antivaccine messages are more liked (*P*<.001) and shared (*P*<.001) but are less viewed (*P*<.001) than provaccine content	N/A	Fear-related themes are present more often in the antivaccine content than in the provaccine one. However, the difference is not statistically significant (*P*>.05)	Antivaccine content is less likely to have been developed by health professionals than provaccine contents (*P*=.002)
Donzelli et al (2018) [43]	Antivaccine messages have more likes, shares, and views (*P*<.001) than provaccine content	Antivaccine content is mainly represented in the categories “People & Blogs” and “No-profit and activism”	N/A	N/A
Ekram et al (2019) [44]	There is no difference between anti- and provaccine content in the number of likes and views (*P*>.05)	N/A	Antivaccine users are more likely to exclude information about vaccine safety (*P*<.05) and efficacy (*P*<.05)	Antivaccine users are more likely to report incorrectly (*P*<.001) or omit information (*P*<.01)
Faasse et al (2016) [45]	Antivaccine comments have fewer likes than provaccine comments (*P*<.001)	Antivaccine content uses more analytical thinking (*P*<.001) and less tentative language (*P*=.055)	Antivaccine users use more words related to health (*P*<.001) and money (*P*=.03)	N/A
Guidry et al (2015) [46]	N/A	Antivaccine users use more narrative than statistical information (*P*<.001)	N/A	27.7% of the antivaccine content mentions conspiracy theories
Harvey et al (2019) [47]	Antivaccine messages get more reactions (*P*<.001) and shares (*P*<.001) than provaccine content	Antivaccine content has stronger emotional appeal (*P*<.001)	Antivaccine content appeals more to fear (*P*<.001) and more often uses themes connecting vaccines with injuries, safety issues, and autism (*P*<.001)	Antivaccine content contains a greater percentage of false statements (*P*<.001) and more often mentions conspiracy theories (*P*<.01)
Kang et al (2017) [48]	N/A	Antivaccine content addresses a broader range of topics compared with provaccine content	Central concepts are thimerosal, mercury, autism, flu shots, and vaccine ingredients	Central concepts are CDC[c], the vaccine industry, mainstream media, doctors, mandatory vaccines, and pharmaceutical companies
Kearney et al (2019) [49]	Antivaccine messages are more liked (*P*<.001) than provaccine ones	Antivaccine content uses more personal narratives than informational ones (*P*<.001). Videos or text with images are more often used rather than text or images alone (*P*<.001)	N/A	Antivaccine content is less likely to have been created by health-related users (*P*<.001)
Luo et al (2019) [50]	N/A	N/A	Antivaccine users more often use words such as “death,” “concern,” “kill,” “injured,” “safety,” “adverse,” “scandal,” and “fraud”	N/A

Study	Hypothesis 1: Antivaccine messages are more liked and shared than provaccine content	Hypothesis 2: Antivaccine users share more user-friendly content than provaccine users	Hypothesis 3: Antivaccine users describe vaccines as harmful for health	Hypothesis 4: Antivaccine users share conspiracy theories or claims that are not scientifically proven
Massey et al (2016) [51]	Negative messages have fewer retweets than provaccine ones do but are still more than neutral ($P<.001$)	N/A	Antivaccine users more often mention side effects ($P<.001$) and are less likely to discuss the protection provided by HPV vaccine ($P<.001$)	N/A
Mitra et al (2016) [52]	N/A	More direct ($P<.05$) and certain language ($P<.05$) is used in antivaccine content	Antivaccine users more often discuss death concerns ($P<.001$)	The “evil government” topic is often discussed ($P<.001$)
Song and Gruzd (2017) [53]	Antivaccine messages have a higher like-to-dislike ratio ($P=.001$)	Antivaccine content is easier to access through YouTube recommendations (centrality measures $P<0.01$). It is more likely to be found in the categories “People & Blogs” and “News & Politics” ($P<.005$)	N/A	N/A
Venkatraman et al (2015) [54]	No difference was noted in terms of likes ($P=.86$) and views ($P=.38$) between anti- and provaccine content	Antivaccine users more often use celebrities in videos ($P=.07$)	Antivaccine users are more likely to use personal stories to indicate the negative effects caused by vaccinations ($P<.001$)	The links provided by antivaccine users less often lead to scientific articles ($P=.01$)
Yiannakoulias et al (2019) [55]	Antivaccine content is more liked than provaccine content is	N/A	Antivaccine content has higher frequency of the words “mercury,” “syringe,” “chemical,” and “toxic”	N/A
Yuan et al (2019) [56]	N/A	Antivaccine users prefer to communicate with users of the same opinion group	N/A	N/A

---

[a] N/A: not applicable.

[b] HPV: human papillomavirus.

[c] CDC: US Centers for Disease Control and Prevention.

Hypothesis 1: Popularity of Content

The authors used different measures to estimate the popularity of antivaccine content. These measures depended on the characteristics of the platforms. Each platform allows researchers to assess user engagement in response to the content. They measured retweets on Twitter [40,51]; likes and shares on Facebook [45,47]; likes on Instagram [39,49]; and likes, dislikes, or shares on YouTube [42-44,53-55]. In addition, YouTube allows for the measurement of audience size by providing information about the number of views, which was also analyzed in a few papers [42-44,54]. In our data set, 12 studies evaluated whether antivaccine content is more popular on social media than provaccine content. Overall, 8 of them found that antivaccine content was more popular, 2 did not find a difference in popularity, and 2 found that provaccine messages were more popular. The popularity of user-generated content depends on the type of platform and can be expressed by likes, reactions, shares, retweets, or views. On Instagram, antivaccine messages had more likes than provaccine messages [39,49]. YouTube videos with negative vaccine sentiments were more appreciated by users (they had more likes, shares, or views) [42,43,53,55]. However, 2 studies did not find statistically significant differences in the number of likes and views between provaccine and antivaccine YouTube videos [44,54]. The analysis of Twitter content did not show a clear link between sentiment and popularity. Blankenship et al [40] found that antivaccine messages had more retweets than provaccine messages, whereas Massey et al [51] found the opposite. The results of the study on Facebook content were similarly ambiguous. Harvey et al [47] found that antivaccine memes had more likes and shares, whereas Faasse et al [45] showed that provaccine Facebook comments had more likes. The results suggest that although antivaccine messages are generally more popular on social media, platform-dependent differences may exist.

Hypothesis 2: User-friendly Content

Antivaccine users present content in a way that tends to be user friendly: it grabs the users’ attention and encourages them to read the posts. Various methods are used to ensure this. The messages contain personal narratives or stories [46,49,54] and are connected with strong emotions and fear [42,47,52]. The language used is also different. Antivaccine articles use phrases such as *mandated vaccines* and *adverse effects*, whereas provaccine articles use *required vaccines* and *side effects* [48]. The language of antivaccine content is direct and certain [52] not tentative [45]. This content is rarely created by health professionals [42,49] but is well suited to the web-based environment as it uses videos or text with images rather than text or images alone [49]. Antivaccine videos also feature celebrities more often than provaccine videos do [54]. On

YouTube, negative videos are mainly presented in lifestyle categories such as *People & Blogs* [43,53]. All results show that antivaccine users create their messages in a way that effectively grabs the attention of their audience.

Hypothesis 3: Vaccines Are Unhealthy or Do Not Work

Antivaccine users often describe vaccinations as harmful to health. They connect them with serious but scientifically unconfirmed injuries, safety issues, and side effects such as autism or even death [44,47,51,54]. They refer to the use of toxic ingredients in vaccines, including mercury and thimerosal [48,55].

Antivaccine messages try to create the impression that vaccines do not protect against diseases. Not only do they present false information about vaccines [47] but they also tend to omit information about the safety and protection provided by vaccines [39,44,51]. All papers indicated that antivaccine users try to create an impression that vaccines are ineffective and dangerous to health.

Hypothesis 4: Vaccines Are a Part of a Conspiracy

In the antivaccine discourse, vaccines are often described as part of conspiracy theories [44,46,47]. Antivaccine concepts are linked to the Centers for Disease Control and Prevention, the Food and Drug Administration, and the pharmaceutical industry hiding the truth about vaccines' side effects or a government using vaccines to obtain control over society or to perform eugenic actions [45,48,50,52]. Antivaccine users view mandatory vaccines as a threat to their civil liberties [46,47]. They believe that as vaccinations carry risks, patients' informed consent is necessary.

Bots' Activity

Using automated tools to spread vaccine-related information can be an efficient and effective way of convincing social media users. A total of 77% of the bots active on Twitter are provaccine. They are hypersocial in retweeting, and they mainly retweet from their opinion group [56]. Research on the activity of bots and trolls in a vaccine debate on Twitter showed that content polluters, which may have been primarily created for marketing, share antivaccine messages at a higher rate than average nonbot users do. Russian trolls post vaccine content at a higher rate than nonbot users do, but they share an equal number of provaccine and antivaccine messages because their purpose is to promote discord in the vaccination debate [41].

Discussion

Principal Findings

Regardless of the social media platform, there are similarities in the characteristics of antivaccine content. Most of the authors found that vaccine-related messages with negative sentiments had a higher number of positive reactions on social media (likes, shares, and retweets). This relationship was particularly evident on YouTube and Instagram, whereas the results from the studies on Twitter and Facebook were inconclusive. The reason for this ambiguity in the case of Facebook may lie in the sampling of the analyzed messages. Harvey et al [47] studied the popularity of memes published on 128 Facebook pages devoted to vaccinations and found that antivaccine messages received more reactions than provaccine messages. Faasse et al [45] obtained the opposite results after studying the comments posted in response to the photo uploaded by Mark Zuckerberg to promote vaccines. However, the percentage of provaccine users constituting Mark Zuckerberg's profile followers may be higher than the percentage of provaccine users in the entire population. On Twitter, Massey et al [51] found that negative messages have fewer retweets than positive ones, but this study was limited to HPV-related tweets. Future research should investigate the popularity of antivaccine content on Twitter and Facebook. The high number of likes on and shares of antivaccine content poses the danger that ordinary users will find this information more easily and consider it to be more reliable than provaccine messages.

We also found that antivaccine users create messages in a user-friendly manner. They publish emotional personal stories using direct language. Their content is not created by health professionals, and they publish more often in the categories under lifestyle than in those connected with medicine or science. Psychological studies have shown that emotional events are remembered more accurately and for a longer time than neutral events [57]. Thus, the emotional communication of antivaccine content may have contributed to the effectiveness of the antivaccine movement. Antivaccine activists often use methods that are commonly used in marketing. Emotional stories attract the attention of neutral users [58], whereas the stirring up of fear of vaccinations leads to the inaction of the audience [59].

The proponents of the antivaccine movement call vaccines dangerous for health and ineffective. The myth-busting research pointing to vaccine safety [60,61] is not stopping the spread of false information. Some people do not trust science, considering it as a tool in the hands of governments and pharmaceutical companies [62]. Many parents believe that the risks of vaccination outweigh its benefits [63]. In countries where parents do not see children dying from infectious diseases, it is easy to think that vaccines, and not the diseases they protect against, are the problem. This phenomenon is perhaps being magnified by the fact that fear-related sources automatically attract attention [64].

Antivaccine messages often contain conspiracy theories. Previous research indicates that a belief in at least one conspiracy theory is common in society [65]. Contrary to popular opinion, conspiracism is not a product of ignorance; it can be explained by the human willingness to believe in the unseen. In a study conducted by Jolley and Douglas [62], participants who were exposed to antivaccine conspiracy theories showed less intention to vaccinate than the control group. Education can help solve this problem because preexisting knowledge about vaccination may protect against the negative effects of exposure to vaccine-related conspiracy theories on the web [66]. Only a simple debunking of misinformation can strengthen its persistence in the community [67] or even increase the harmful activity of conspiracists [68]. Clarifying parental concerns and involving parents in decisions regarding their child's vaccination can reduce beliefs in conspiracies [69].

The analyzed papers show that bots on social media spread not only antivaccine messages but also provaccine messages. The literature differentiates between benign and malicious bots [70]. Benign bots respond automatically, aggregate content, and perform other useful actions. However, malicious bots are designed with a purpose to harm. Their task is to manipulate, mislead, and exploit to influence social media discourse. Public health authorities should not only monitor social media, detect negative bots, and fight the spread of the antivaccine content, but they should also use benign bots to communicate with the public and dispel doubts about vaccinations.

The results of the included papers are generally consistent with those of previous research examining antivaccine website content [2,6,9,10]. Similar to websites, antivaccine users of social media raise and discuss concerns about the safety and effectiveness of vaccines. They describe vaccines as harmful for health, present scientifically unconfirmed claims, and hide information about the safety and effectiveness of vaccines to strengthen their messages. This behavior can be explained by a well-known cognitive phenomenon called *motivated reasoning*, which is defined as a tendency to find arguments in favor of the conclusions we believe in, rather than those supporting what we do not want to believe in [71]. Both antivaccine websites and antivaccine social media users presented two arguments concerning the government. They believe that the government is cooperating with pharmaceutical corporations to impose universal immunization programs and is allowing these companies to profit. The second argument does not suggest evil government plans but points to the restriction of freedom of choice because of mandatory vaccines. Unlike the content of antivaccine websites, negative posts on social media do not promote alternative health practices; they simply discourage vaccinations. This may be because short and eye-catching content is preferred on social media, and this does not allow for the explanation or discussion of complex issues.

We found that most included studies were based on Twitter data. This is surprising because Facebook, YouTube, and Instagram have many more active users [72]. The reason for this disproportionate attention may lie in the simplicity of gathering data from Twitter. Twitter enables the downloading of thousands of posts using its official application programming interface [73], whereas Facebook and Instagram closed their application programming interfaces in 2018, thus preventing the automatic downloading of publicly available data from these platforms to protect users' data against inappropriate use [74]. This was in response to the Cambridge Analytica data misuse scandal [75]. Moreover, Facebook allows for the creation of closed and private groups whose content is not available to scientists. YouTube provides researchers with easy access to the content by providing automatically generated transcripts of videos, thus facilitating text analysis. Since 2019, Pinterest has been hosting vaccine-related information only from reliable sources to halt the spread of vaccine misinformation [76]; therefore, further research into the antivaccine user's activity on this platform since 2019 is pointless.

Web-based platforms differ in terms of how easily the antivaccine content is spread through social media. Facebook and Instagram have taken steps to stop the spread of vaccine misinformation by making it less prominent in search results; however, such misinformation has not been completely removed from these platforms [77]. Recent research shows that since 2016, interactions with content containing misinformation have reduced on Facebook but have continued to increase on Twitter [78]. This suggests that misinformation on Twitter can become a bigger problem than on Facebook. During the COVID-19 pandemic in 2020, Twitter started removing accounts spreading health conspiracy theories [79] and tried to redirect users searching for COVID-19–related information to official government sources [80].

YouTube facilitates the spread of misinformation to millions of viewers. It has been found that 27.5% of the most-viewed YouTube videos related to COVID-19 contained misinformation, and these reached over 62 million viewers worldwide [81]. About 17 million people subscribe to antivaccine accounts on YouTube [82]. As YouTube content is more difficult to create, it is probably less affected by bots sharing video content. However, bots can spread disinformation on YouTube through comments [83].

In the papers studied, the HPV vaccine was the second most common topic, after the topic of vaccines in general. This topic is popular in the discourse on antivaccine movements. The HPV vaccine has some specific features that make it vulnerable to theories that discourage vaccinations. First, in many countries, this vaccine is not mandatory; therefore, it can be more effectively discouraged. Second, the minimum age for receiving the first dose is 9 years; thus, often both parents' consent and the preteen or teenager's acceptance are required. Finally, apart from the ordinary antivaccination supporters, the HPV vaccine has other opponents. Conservative religious groups suggest that protection against sexually transmitted diseases encourages licentious teen sex. As a result, religious young women are undervaccinated and underinformed about the HPV vaccine [84].

Future Research Propositions

Research conducted in languages other than English is limited. In our data set, only 2 articles analyzed non-English messages, which were in Italian [42,43]. We propose that a multilingual comparative study be undertaken to explore the similarities and differences in the vaccine-related discourse on social media between countries.

Another issue is the classification of messages as provaccine or antivaccine. Usually, this task is performed manually using codebooks. As this method is not scalable, we need to employ an automatic approach. Popularly, some of the messages are hand-labeled using a codebook, and machine learning models are applied to label the rest of the messages. We suggest that a universal codebook or dictionary be designed to assess whether a message has antivaccine sentiment. Such a tool would enhance the comparability of the research results.

Instagram is a social media platform that is still gaining popularity not only among users but also among scientists, as shown by the 2 articles from 2019 [39,49] that we included in our systematic review. However, previous Instagram studies have only examined opinions regarding HPV vaccines.

Analyzing data from private or secret Facebook groups [85] is important, as they often contain valuable information about users' opinions and attitudes toward vaccinations. Although messages posted in such groups potentially contain higher quality information than publicly available data, they are not easy to obtain.

It should also be mentioned that the articles analyzed picture antivaccine messages related to specific words, phrases, and themes. Further research on these topics can result in the creation of effective tools to automatically detect fake news. This can help suppress vaccine hesitancy, which is connected to the spread of vaccine misinformation on social media [86].

Limitations

First, we restricted the analysis to studies published between 2015 and 2019. As a result, some papers describing antivaccine messages on social media may not have been included. Second, we studied only articles written in English. Thus, we may have omitted articles published in other languages that analyzed non-English messages on social media.

Conclusions

Antivaccine users create content that gains more positive reactions (likes, shares, retweets) on social media than provaccine messages do. Their messages are user friendly and well suited to the needs of users on social media platforms. Antivaccine users try to discourage vaccination using a few main arguments. They describe vaccines as harmful, highlight their side effects, and undermine the effectiveness of the protection they offer. To support these statements, they use false information and conspiracy theories, and gloss over or omit the data about vaccine benefits.

Public health authorities should continuously monitor social media to find new antivaccine arguments quickly and, based on that, design information campaigns targeting health professionals and ordinary users who are at a risk of being misinformed. Social media platforms have a big responsibility because they give millions of users access to misinformation. Knowledge of the characteristics of antivaccine content can help in the creation of tools that automatically tag false information. A positive trend in recent years is that social media platforms have attempted to stop the spread of vaccination misinformation.

Conflicts of Interest

None declared.

References

1. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet: caveant lector et viewor--Let the reader and viewer beware. J Am Med Assoc 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
2. Wolfe RM, Sharp LK, Lipsky MS. Content and design attributes of antivaccination web sites. J Am Med Assoc 2002 Jun 26;287(24):3245-3248. [doi: 10.1001/jama.287.24.3245] [Medline: 12076221]
3. Wolfe RM, Sharp LK. Vaccination or immunization? The impact of search terms on the internet. J Health Commun 2005 Sep;10(6):537-551. [doi: 10.1080/10810730500228847] [Medline: 16203632]
4. Zimmerman RK, Wolfe RM, Fox DE, Fox JR, Nowalk MP, Troy JA, et al. Vaccine criticism on the World Wide Web. J Med Internet Res 2005 Jun 29;7(2):e17 [FREE Full text] [doi: 10.2196/jmir.7.2.e17] [Medline: 15998608]
5. Bean SJ. Emerging and continuing trends in vaccine opposition website content. Vaccine 2011 Feb 24;29(10):1874-1880. [doi: 10.1016/j.vaccine.2011.01.003] [Medline: 21238571]
6. Davies P, Chapman S, Leask J. Antivaccination activists on the world wide web. Arch Dis Child 2002 Jul;87(1):22-25 [FREE Full text] [doi: 10.1136/adc.87.1.22] [Medline: 12089115]
7. Glanz JM, Wagner NM, Narwaney KJ, Kraus CR, Shoup JA, Xu S, et al. Web-based social media intervention to increase vaccine acceptance: a randomized controlled trial. Pediatrics 2017 Dec;140(6):- [FREE Full text] [doi: 10.1542/peds.2017-1117] [Medline: 29109107]
8. Evrony A, Caplan A. The overlooked dangers of anti-vaccination groups' social media presence. Hum Vaccin Immunother 2017 Jun 3;13(6):1-2 [FREE Full text] [doi: 10.1080/21645515.2017.1283467] [Medline: 28406737]
9. 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]
10. 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]
11. Amin AB, Bednarczyk RA, Ray CE, Melchiori KJ, Graham J, Huntsinger JR, et al. Association of moral values with vaccine hesitancy. Nat Hum Behav 2017 Dec;1(12):873-880. [doi: 10.1038/s41562-017-0256-5] [Medline: 31024188]
12. Benecke O, DeYoung SE. Anti-vaccine decision-making and measles resurgence in the United States. Glob Pediatr Health 2019;6:2333794X19862949 [FREE Full text] [doi: 10.1177/2333794X19862949] [Medline: 31384629]
13. Diethelm P, McKee M. Denialism: what is it and how should scientists respond? Eur J Public Health 2009 Jan;19(1):2-4. [doi: 10.1093/eurpub/ckn139] [Medline: 19158101]

14. Vaccine Hesitancy. European Centre for Disease Prevention and Control. URL: https://www.ecdc.europa.eu/en/immunisation-vaccines/vaccine-hesitancy [accessed 2021-05-20]
15. 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]
16. Dubé E, Gagnon D, MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Strategies intended to address vaccine hesitancy: review of published reviews. Vaccine 2015 Aug 14;33(34):4191-4203 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.041] [Medline: 25896385]
17. Dubé E, Gagnon D, Ouakki M, Bettinger JA, Guay M, Halperin S, Canadian Immunization Research Network. Understanding vaccine hesitancy in Canada: results of a consultation study by the Canadian immunization research network. PLoS One 2016;11(6):e0156118 [FREE Full text] [doi: 10.1371/journal.pone.0156118] [Medline: 27257809]
18. Report of the Sage Working Group on Vaccine Hesitancy. World Health Organization. 2014. URL: https://www.who.int/immunization/sage/meetings/2014/october/1_Report_WORKING_GROUP_vaccine_hesitancy_final.pdf [accessed 2021-05-27]
19. Mason BW, Donnelly PD. Impact of a local newspaper campaign on the uptake of the measles mumps and rubella vaccine. J Epidemiol Community Health 2000 Jun;54(6):473-474 [FREE Full text] [doi: 10.1136/jech.54.6.473] [Medline: 10818125]
20. Smith A, Yarwood J, Salisbury DM. Tracking mothers' attitudes to MMR immunisation 1996-2006. Vaccine 2007 May 16;25(20):3996-4002. [doi: 10.1016/j.vaccine.2007.02.071] [Medline: 17395344]
21. 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]
22. Betsch C, Renkewitz F, Betsch T, Ulshöfer C. The influence of vaccine-critical websites on perceiving vaccination risks. J Health Psychol 2010 Apr;15(3):446-455. [doi: 10.1177/1359105309353647] [Medline: 20348365]
23. Dubé E, Bettinger JA, Halperin B, Bradet R, Lavoie F, Sauvageau C, et al. Determinants of parents' decision to vaccinate their children against rotavirus: results of a longitudinal study. Health Educ Res 2012 Dec;27(6):1069-1080. [doi: 10.1093/her/cys088] [Medline: 22907535]
24. Smith PJ, Humiston SG, Marcuse EK, Zhao Z, Dorell CG, Howes C, et al. Parental delay or refusal of vaccine doses, childhood vaccination coverage at 24 months of age, and the Health Belief Model. Public Health Rep 2011;126(Suppl 2):135-146 [FREE Full text] [doi: 10.1177/00333549111260S215] [Medline: 21812176]
25. Witteman HO, Zikmund-Fisher BJ. The defining characteristics of web 2.0 and their potential influence in the online vaccination debate. Vaccine 2012 May 28;30(25):3734-3740. [doi: 10.1016/j.vaccine.2011.12.039] [Medline: 22178516]
26. Young SD. Social media as a new vital sign: commentary. J Med Internet Res 2018 Apr 30;20(4):e161 [FREE Full text] [doi: 10.2196/jmir.8563] [Medline: 29712631]
27. Charalambous A. Social media and health policy. Asia Pac J Oncol Nurs 2019;6(1):24-27 [FREE Full text] [doi: 10.4103/apjon.apjon_60_18] [Medline: 30599012]
28. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine 2018 Jun 14;36(25):3606-3612. [doi: 10.1016/j.vaccine.2018.05.040] [Medline: 29773322]
29. Ward JK, Peretti-Watel P, Verger P. Vaccine criticism on the internet: propositions for future research. Hum Vaccin Immunother 2016 Jul 2;12(7):1924-1929 [FREE Full text] [doi: 10.1080/21645515.2016.1146430] [Medline: 26900646]
30. Langenfeld SJ, Batra R. How can social media get us in trouble? Clin Colon Rectal Surg 2017 Sep;30(4):264-269 [FREE Full text] [doi: 10.1055/s-0037-1604255] [Medline: 28924400]
31. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
32. Moving the Needle: Promoting Vaccination Uptake across the Life Course. Royal Society for Public Health UK. URL: https://www.rsph.org.uk/our-work/policy/vaccinations/moving-the-needle-promoting-vaccination-uptake-across-the-life-course.html [accessed 2020-05-15]
33. Dredze M, Broniatowski DA, Smith MC, Hilyard KM. Understanding vaccine refusal: why we need social media now. Am J Prev Med 2016 Apr;50(4):550-552 [FREE Full text] [doi: 10.1016/j.amepre.2015.10.002] [Medline: 26655067]
34. 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]
35. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194 [FREE Full text] [doi: 10.1177/1460458213512220] [Medline: 24670899]
36. Catalan-Matamoros D, Peñafiel-Saiz C. How is communication of vaccines in traditional media: a systematic review. Perspect Public Health 2019 Jan 7;139(1):34-43. [doi: 10.1177/1757913918780142] [Medline: 29877122]
37. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
38. Duggan M, Lenhart A, Dimock M, Gross E, Tomlin R, Kohut A. Social Media Update. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/01/09/social-media-update-2014/ [accessed 2021-05-25]

39. 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]
40. Blankenship EB, Goff ME, Yin J, Tse ZT, Fu K, Liang H, et al. Sentiment, contents, and retweets: a study of two vaccine-related twitter datasets. Perm J 2018;22:17-138 [FREE Full text] [doi: 10.7812/TPP/17-138] [Medline: 29911966]
41. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
42. Covolo L, Ceretti E, Passeri C, Boletti M, Gelatti U. What arguments on vaccinations run through YouTube videos in Italy? A content analysis. Hum Vaccin Immunother 2017 Jul 3;13(7):1693-1699 [FREE Full text] [doi: 10.1080/21645515.2017.1306159] [Medline: 28362544]
43. Donzelli G, Palomba G, Federigi I, Aquino F, Cioni L, Verani M, et al. Misinformation on vaccination: a quantitative analysis of YouTube videos. Hum Vaccin Immunother 2018 Jul 3;14(7):1654-1659 [FREE Full text] [doi: 10.1080/21645515.2018.1454572] [Medline: 29553872]
44. 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]
45. Faasse K, Chatman CJ, Martin LR. A comparison of language use in pro- and anti-vaccination comments in response to a high profile Facebook post. Vaccine 2016 Nov 11;34(47):5808-5814. [doi: 10.1016/j.vaccine.2016.09.029] [Medline: 27707558]
46. Guidry JP, Carlyle K, Messner M, Jin Y. On pins and needles: how vaccines are portrayed on Pinterest. Vaccine 2015 Sep 22;33(39):5051-5056. [doi: 10.1016/j.vaccine.2015.08.064] [Medline: 26319742]
47. Harvey AM, Thompson S, Lac A, Coolidge FL. Fear and derision: a quantitative content analysis of provaccine and antivaccine internet memes. Health Educ Behav 2019 Dec;46(6):1012-1023. [doi: 10.1177/1090198119866886] [Medline: 31789076]
48. 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]
49. Kearney MD, Selvan P, Hauer MK, Leader AE, Massey PM. Characterizing HPV vaccine sentiments and content on Instagram. Health Educ Behav 2019 Dec;46(2_suppl):37-48. [doi: 10.1177/1090198119859412] [Medline: 31742459]
50. Luo X, Zimet G, Shah S. A natural language processing framework to analyse the opinions on HPV vaccination reflected in twitter over 10 years (2008 - 2017). Hum Vaccin Immunother 2019;15(7-8):1496-1504 [FREE Full text] [doi: 10.1080/21645515.2019.1627821] [Medline: 31194609]
51. 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 5;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
52. Mitra T, Counts S, Pennebaker JW. Understanding Anti-Vaccination Attitudes in Social Media. AAAI. 2016. URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/download/13073/12747 [accessed 2021-05-25]
53. Song M, Gruzd A. Examining Sentiments and Popularity of Pro-and Anti-vaccination Videos on Youtube. In: International Conference Proceeding Series. 2017 Presented at: ACM'17; April 12-17, 2017; New York, USA. [doi: 10.1145/3097286.3097303]
54. Venkatraman A, Garg N, Kumar N. Greater freedom of speech on web 2.0 correlates with dominance of views linking vaccines to autism. Vaccine 2015 Mar 17;33(12):1422-1425. [doi: 10.1016/j.vaccine.2015.01.078] [Medline: 25665960]
55. Yiannakoulias N, Slavik CE, Chase M. Expressions of pro- and anti-vaccine sentiment on YouTube. Vaccine 2019 Apr 3;37(15):2057-2064. [doi: 10.1016/j.vaccine.2019.03.001] [Medline: 30862365]
56. Yuan X, Schuchard RJ, Crooks AT. Examining emergent communities and social bots within the polarized online vaccination debate in Twitter. Soc Med Soc 2019 Sep 4;5(3):205630511986546. [doi: 10.1177/2056305119865465]
57. Tyng CM, Amin HU, Saad MN, Malik AS. The influences of emotion on learning and memory. Front Psychol 2017;8:1454 [FREE Full text] [doi: 10.3389/fpsyg.2017.01454] [Medline: 28883804]
58. Bagozzi RP, Gopinath M, Nyer PU. The role of emotions in marketing. J Acad Mark Sci 1999 Apr 1;27(2):184-206. [doi: 10.1177/0092070399272005]
59. Brennan L, Binney W. Fear, guilt, and shame appeals in social marketing. J Bus Res 2010 Feb;63(2):140-146. [doi: 10.1016/j.jbusres.2009.02.006]
60. Jain A, Marshall J, Buikema A, Bancroft T, Kelly JP, Newschaffer CJ. Autism occurrence by MMR vaccine status among US children with older siblings with and without autism. J Am Med Assoc 2015 Apr 21;313(15):1534-1540. [doi: 10.1001/jama.2015.3077] [Medline: 25898051]
61. Ray P, Hayward J, Michelson D, Lewis E, Schwalbe J, Black S, Vaccine Safety Datalink Group. Encephalopathy after whole-cell pertussis or measles vaccination: lack of evidence for a causal association in a retrospective case-control study. Pediatr Infect Dis J 2006 Sep;25(9):768-773. [doi: 10.1097/01.inf.0000234067.84848.e1] [Medline: 16940831]
62. 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]

63. Karafillakis E, Larson HJ, ADVANCE consortium. The benefit of the doubt or doubts over benefits? A systematic literature review of perceived risks of vaccines in European populations. Vaccine 2017 Sep 5;35(37):4840-4850 [FREE Full text] [doi: 10.1016/j.vaccine.2017.07.061] [Medline: 28760616]
64. Schmidt LJ, Belopolsky AV, Theeuwes J. Potential threat attracts attention and interferes with voluntary saccades. Emotion 2015 Jun;15(3):329-338. [doi: 10.1037/emo0000041] [Medline: 25527964]
65. Oliver JE, Wood TJ. Conspiracy theories and the paranoid style(s) of mass opinion. Am J Pol Sci 2014 Mar 5;58(4):952-966. [doi: 10.1111/ajps.12084]
66. Chen L, Zhang Y, Young R, Wu X, Zhu G. Effects of vaccine-related conspiracy theories on Chinese young adults' perceptions of the HPV vaccine: an experimental study. Health Commun 2020 Apr 20:1-11. [doi: 10.1080/10410236.2020.1751384] [Medline: 32312084]
67. Chan MS, Jones CR, Jamieson K, Albarracín D. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. Psychol Sci 2017 Nov;28(11):1531-1546 [FREE Full text] [doi: 10.1177/0956797617714579] [Medline: 28895452]
68. Zollo F, Bessi A, Del Vicario M, Scala A, Caldarelli G, Shekhtman L, et al. Debunking in a world of tribes. PLoS One 2017;12(7):e0181821 [FREE Full text] [doi: 10.1371/journal.pone.0181821] [Medline: 28742163]
69. van Prooijen J, Acker M. The influence of control on belief in conspiracy theories: conceptual and applied extensions. Appl Cognit Psychol 2015 Aug 10;29(5):753-761. [doi: 10.1002/acp.3161]
70. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun ACM 2016 Jun 24;59(7):96-104. [doi: 10.1145/2818717]
71. Kunda Z. The case for motivated reasoning. Psychol Bull 1990 Nov;108(3):480-498. [doi: 10.1037/0033-2909.108.3.480] [Medline: 2270237]
72. Ortiz-Ospina E. The Rise of Social Media. Our World in Data. 2019. URL: https://ourworldindata.org/rise-of-social-media [accessed 2020-04-10]
73. Campan A, Atnafu T, Truta T, Nolan J. Is Data Collection through Twitter Streaming API Useful for Academic Research? In: International Conference on Big Data. 2018 Presented at: IEEE'18; May 6-9, 2018; Seattle. Washington p. 3638-3643. [doi: 10.1109/bigdata.2018.8621898]
74. Freelon D. Computational research in the post-API age. Pol Commun 2018 Oct 25;35(4):665-668. [doi: 10.1080/10584609.2018.1477506]
75. Isaak J, Hanna MJ. User data privacy: Facebook, Cambridge analytica, and privacy protection. Computer 2018 Aug;51(8):56-59. [doi: 10.1109/mc.2018.3191268]
76. Ozoma I. Bringing Authoritative Vaccine Results to Pinterest Search. Pinterest Newsroom. 2019. URL: https://newsroom.pinterest.com/en/post/bringing-authoritative-vaccine-results-to-pinterest-search [accessed 2020-04-14]
77. Matsakis L. Facebook Will Crack Down on Anti-Vaccine Content. Wired. 2019. URL: https://www.wired.com/story/facebook-anti-vaccine-crack-down/ [accessed 2021-05-21]
78. Allcott H, Gentzkow M, Yu C. Trends in the diffusion of misinformation on social media. Res Polit 2019;6(2):-. [doi: 10.3386/w25500]
79. Hawkins D. Twitter Bans Zero Hedge Account After It Doxxed a Chinese Researcher Over Coronavirus. Washington Post. URL: https://www.washingtonpost.com/technology/2020/02/01/twitter-zero-hedge-coronavirus/ [accessed 2020-05-01]
80. Toh M. Facebook, Google and Twitter Crack Down on Fake Coronavirus 'Cures' And Other Misinformation. CNN Business Edition. URL: https://edition.cnn.com/2020/01/31/tech/facebook-twitter-google-coronavirus-misinformation/index.html [accessed 2020-04-12]
81. Li HO, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health 2020 May;5(5):- [FREE Full text] [doi: 10.1136/bmjgh-2020-002604] [Medline: 32409327]
82. Burki T. The online anti-vaccine movement in the age of COVID-19. Lancet Digit Health 2020 Oct;2(10):e504-e505 [FREE Full text] [doi: 10.1016/S2589-7500(20)30227-2] [Medline: 32984795]
83. Hussain M, Tokdemir S, Agarwal N, Al-Khateeb S. Analyzing Disinformation and Crowd Manipulation Tactics on Youtube. In: Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2018 Presented at: ASONAM'18; August 28-31, 2018; Barcelona, Spain p. 1092-1095. [doi: 10.1109/asonam.2018.8508766]
84. Bodson J, Wilson A, Warner EL, Kepka D. Religion and HPV vaccine-related awareness, knowledge, and receipt among insured women aged 18-26 in Utah. PLoS One 2017;12(8):e0183725 [FREE Full text] [doi: 10.1371/journal.pone.0183725] [Medline: 28841681]
85. Franz D, Marsh HE, Chen JI, Teo AR. Using Facebook for qualitative research: a brief primer. J Med Internet Res 2019 Aug 13;21(8):e13544 [FREE Full text] [doi: 10.2196/13544] [Medline: 31411143]
86. 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]

Abbreviations

HPV: human papillomavirus

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

Edited by G Eysenbach; submitted 25.09.20; peer-reviewed by N Yiannakoulias, A Schmidt; comments to author 29.10.20; revised version received 20.12.20; accepted 14.04.21; published 04.06.21.

Please cite as:
Wawrzuta D, Jaworski M, Gotlib J, Panczyk M
Characteristics of Antivaccine Messages on Social Media: Systematic Review
J Med Internet Res 2021;23(6):e24564
URL: https://www.jmir.org/2021/6/e24564
doi:10.2196/24564
PMID:34085943

Review

Evaluation of the Effectiveness of Digital Technology Interventions to Reduce Loneliness in Older Adults: Systematic Review and Meta-analysis

Syed Ghulam Sarwar Shah[1,2], MBBS, MA, MSc, PhD; David Nogueras[3], MBA; Hugo Cornelis van Woerden[4,5,6], MBChB, MPH, PhD, FFPH; Vasiliki Kiparoglou[1,7], BSc, MSc, MBA, PhD

[1]NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

[2]Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom

[3]EvZein Limited, Holley Crescent, Headington, Oxford, United Kingdom

[4]Public Health Agency Northern Ireland, Belfast, United Kingdom

[5]Division of Rural Health and Wellbeing, University of the Highlands and Islands, Inverness, United Kingdom

[6]Institute of Nursing and Health Research, Ulster University, Belfast, United Kingdom

[7]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Syed Ghulam Sarwar Shah, MBBS, MA, MSc, PhD
NIHR Oxford Biomedical Research Centre
Oxford University Hospitals NHS Foundation Trust
John Radcliffe Hospital
Oxford, OX3 9DU
United Kingdom
Phone: 44 1865221262
Email: sarwar.shah@ouh.nhs.uk

Abstract

Background: Loneliness is a serious public health issue, and its burden is increasing in many countries. Loneliness affects social, physical, and mental health, and it is associated with multimorbidity and premature mortality. In addition to social interventions, a range of digital technology interventions (DTIs) are being used to tackle loneliness. However, there is limited evidence on the effectiveness of DTIs in reducing loneliness, especially in adults. The effectiveness of DTIs in reducing loneliness needs to be systematically assessed.

Objective: The objective of this study is to assess the effectiveness of DTIs in reducing loneliness in older adults.

Methods: We conducted electronic searches in PubMed, MEDLINE, CINAHL, Embase, and Web of Science for empirical studies published in English from January 1, 2010, to July 31, 2019. The study selection criteria included interventional studies that used any type of DTIs to reduce loneliness in adults (aged ≥18 years) with a minimum intervention duration of 3 months and follow-up measurements at least 3 months after the intervention. Two researchers independently screened articles and extracted data using the PICO (participant, intervention, comparator, and outcome) framework. The primary outcome measure was loneliness. Loneliness scores in both the intervention and control groups at baseline and at follow-up at 3, 4, 6, and 12 months after the intervention were extracted. Data were analyzed via narrative synthesis and meta-analysis using RevMan (The Cochrane Collaboration) software.

Results: A total of 6 studies were selected from 4939 screened articles. These studies included 1 before and after study and 5 clinical trials (4 randomized clinical trials and 1 quasi-experimental study). All of these studies enrolled a total of 646 participants (men: n=154, 23.8%; women: n=427, 66.1%; no gender information: n=65, 10.1%) with an average age of 73-78 years (SD 6-11). Five clinical trials were included in the meta-analysis, and by using the random effects model, standardized mean differences (SMDs) were calculated for each trial and pooled across studies at the 3-, 4-, and 6-month follow-ups. The overall effect estimates showed no statistically significant difference in the effectiveness of DTIs compared with that of usual care or non-DTIs at follow-up at 3 months (SMD 0.02; 95% CI −0.36 to 0.40; *P*=.92), 4 months (SMD −1.11; 95% CI −2.60 to 0.38; *P*=.14), and 6 months (SMD −0.11; 95% CI −0.54 to 0.32; *P*=.61). The quality of evidence was very low to moderate in these trials.

Conclusions: Our meta-analysis shows no evidence supporting the effectiveness of DTIs in reducing loneliness in older adults. Future research may consider randomized controlled trials with larger sample sizes and longer durations for both the interventions and follow-ups.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-032455

(J Med Internet Res 2021;23(6):e24712) doi:10.2196/24712

KEYWORDS

loneliness; older people; digital technology; effectiveness; efficacy; evidence; systematic review; meta-analysis

Introduction

Background

Loneliness is a multifaceted public health problem [1]. The burden of loneliness is high in some countries [2-9], and it is increasing in many other countries [10]. Loneliness is expected to rise because of lockdowns, quarantine, self-isolation, and social distancing measures that are being enforced in several countries to tackle the COVID-19 pandemic [11,12]. Therefore, tackling loneliness is imperative, and digital technology could play a major role in addressing loneliness [13].

Loneliness refers to an individual's subjective feelings of a perceived discrepancy between actual and desired social relationships [14,15]. Although loneliness affects people of all ages [15,16], older, younger, and vulnerable people are affected more by it [7,17,18]. Risk factors of loneliness include demographic characteristics, social factors, and physical environments [17-19]. Loneliness enhances the risk of poor physical and mental health [14,20-23], dementia [24], premature mortality, and all-cause mortality [21], particularly in older adults [23]. In addition, the implications of loneliness include the high costs of health and well-being (eg, between £6429.00 [US $8074.80] and £9616.00 [US $12,077.70] per person per year in the United Kingdom) [25] as well as lost work days and productivity (eg, costing up to £2.5 billion [US $3.14 billion] per annum for employers in the United Kingdom) [26]. Therefore, it is imperative to tackle loneliness.

Loneliness is being addressed through a range of social [27] and technological interventions [28]. The latter type of interventions includes numerous and diverse types of digital apps, web-based social networking tools, sensors, and robots [29]. Although these tools use digital technology, they are heterogeneous in many aspects, including the means they provide to socially connect; the purposes for which they are used; the ways and methods of their application; the frequency of their use; and their users, who differ from each other in many traits such as demographic, social, and economic characteristics, and some may have physical and mental limitations. Therefore, these digital technology tools need to be systematically evaluated for their effectiveness in tackling loneliness.

Several published reviews have reported that digital technology interventions (DTIs) are effective in reducing loneliness [30-34]. However, some of these studies are weak and have a high risk of bias [35], and other studies have used a few selected technological interventions and covered literature published over a short span, such as the 3-year period from January 2010 to January 2013 [31].

There is limited evidence on the effectiveness of DTIs for loneliness [36], and there are calls for further research [32,33] to assess and identify the latest DTIs that are effective in reducing loneliness [34,36]. In addition, evaluation of the latest evidence on the effectiveness of DTIs in reducing loneliness is imperative from the perspectives of patients and their families and other stakeholders such as health and social care providers and health insurers [37].

Study Objectives

The primary objective of this study is to assess the effectiveness of DTIs in reducing loneliness in adults. The secondary objective is to identify DTIs that are used to reduce loneliness in adults.

Review Questions

The main research question was "Are DTIs effective for reducing loneliness in adults?" The secondary question was "What DTIs are used for reducing loneliness in adults?"

Outcome Measures

The main outcome measure was loneliness. We extracted data on loneliness measured at both the baseline (before the intervention) and follow-ups (at least 3 months after the intervention) for the intervention groups and control groups, if any.

Methods

Study Design, Conduct, and Reporting

We undertook a systematic review and meta-analysis as suggested in the Cochrane Methods for Systematic Reviews of Interventions [38]. We have reported the findings in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [39].

Protocol Registration and Publication

We registered this systematic review and meta-analysis with the PROSPERO database on June 10, 2019 (registration ID: CRD42019131524) [40], and we published our protocol [37] before undertaking this study.

Patient and Public Involvement

A patient and public manager affiliated with our research center reviewed the study protocol and provided suggestions that were incorporated into the protocol. We had no access to any patient diagnosed with loneliness; therefore, we could not include any patients or members of the public in the design and conduct of the study. However, the findings of this study will be disseminated as an open access publication that will be freely available to patients and everyone else globally.

Definition of DTI

We define the term DTI as an intervention that applies digital technology, that is, the technology, equipment, and apps that process information in the form of numeric codes, usually a binary code [41].

Eligibility Criteria

We selected studies that met our predefined eligibility criteria [37]. Study designs included interventional studies (randomized and nonrandomized) that investigated the effects of DTIs on loneliness. We included a range of DTIs, that is, computers, computer tablets, iPads, internet, web-based videos, communication, chatting, social groups, meetings, conferences and messages, sensors, social robots, smart mobile phones, social media tools, and the World Wide Web. We set 3 months as the minimum intervention duration and follow-up period. The research participants were adults, both male and female, aged 18 years or more. We included different settings, that is, residential dwellings, including private residences and care or nursing homes or centers in any country. The studies were limited to journal articles in English published from January 1, 2010, to July 31, 2019.

Information Sources and Keywords

We electronically searched PubMed, MEDLINE, CINAHL, Embase, and Web of Science and covered the publication period from January 1, 2010, to July 31, 2019. We used an a priori list of keywords prepared in our preliminary literature searches [37]. The keywords were of 2 categories: medical condition or problem (ie, *loneliness*, *lonely*, *isolation*, *aloneness*, *disconnect**, *solitude*, *singleness**, *lonesomeness*, *solitariness*, and *remoteness*) and intervention or technology (ie, *digital*, *technolog**, *sensor**, *robot**, *internet*, *social media*, **phone**, *online*, *iPad**, *tablet**, *computer**, *electronic*, *web*, *video*, and *videoconference*), as reported in our published protocol [37].

Literature Searches

First, we searched the keywords in the *subject headings* such as MeSH major terms in PubMed or equivalent terms in other databases (for detailed search history, see Multimedia Appendix 1). Thereafter, we searched for keywords in the *title* and *abstract* fields in the selected databases using 3 Boolean operators: “OR,” “AND,” and “NOT.” In addition, we hand searched the reference lists of the shortlisted articles. We wrote emails to the authors of 2 studies requesting for full copies of their research articles [42,43], which were gratefully emailed to us. We contacted the authors of 2 further studies for missing or additional data [44,45]. We had a good response from the authors of both studies, and data were thankfully provided for 1 study only [45]. We sought support from an expert librarian at our library for running literature searches.

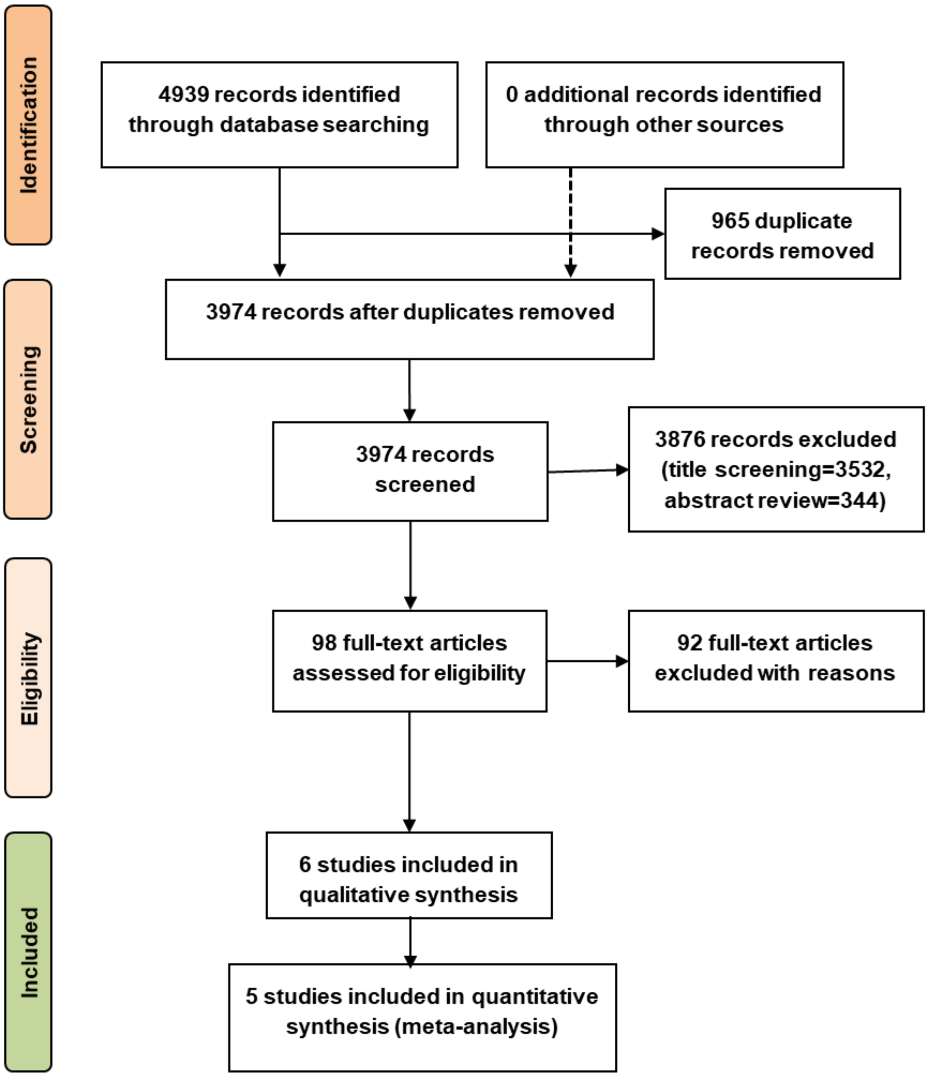
Study Selection

Literature searches retrieved 4939 articles, of which 965 duplicate articles were removed (Figure 1). Two researchers (SGSS and DN) independently screened the remaining articles (n=3974) by title, which was followed by reading the abstracts of 442 articles (Figure 1). This screening process led to the exclusion of 3876 articles and identification of 98 articles for full-text review. Three researchers (SGSS, DN, and VK) independently read the full texts of these 98 articles.

When recommendations differed between reviewers at the title, abstract, and full-text review stages, another reviewer (HCvW) reviewed these articles, and his recommendations to either include or exclude an article were final.

Finally, 92 articles were excluded, and the remaining 6 articles were included in the data extraction (Figure 1). All these 6 studies were included in the narrative synthesis, whereas 5 studies—all clinical trials involving an intervention group and a control group—were included in the meta-analysis (Figure 1). One study with a pre- and postintervention design involving only the intervention group was excluded from the meta-analysis.

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



Data Collection Process

For data collection, we used an a priori data extraction template (Tables 1 and 2), which comprised several columns: authors, year, and country of study; study aim or objectives; research design; settings; participants' characteristics (age, gender, and ethnicity); health or medical condition; sampling method and sample size; participant attrition (numbers and percentages); research methods and data collection tools; interventions (eg, type and tool of digital technology); comparators (eg, alternative intervention, placebo, or care as usual); intervention duration (weeks or months); measurement stages (eg, baseline and follow-up: weeks or months after the baseline); outcomes, result, and findings (eg, loneliness scores, including statistics; eg, mean values, SDs, SEs, and CIs); and study authors' conclusions [37].

Table 1. Characteristics of included studies, participants, sampling methods and sizes, and data collection tools.

Study, country	Quality of evidence[a] (reviewers' assessment)	Research design	Settings	Participants			Main health or medical conditions investigated	Sampling method	Sample size			Participant attrition	Research methods or data collection tools
				Age (years)	Gender	Ethnicity			Total	Intervention group	Control group		Loneliness scale used
Tsai et al (2010) [46], Taiwan	Medium	Quasi-experimental study (NRCT[b])	Nursing home	Baseline: experimental group: average age 74.2 (SD 10.18); control group: average age 78.48 (SD 6.75)	Male=24 (experimental group=10; control group=14); female=33 (experimental group=14; control group=19)	Not reported (probably all Taiwanese or Chinese)	Loneliness and depression	Purposive	57 baseline; 49 end of study	24 baseline; 21 follow-up	33 baseline; 28 follow-up	8 (5 from control group and 3 from experimental group); attrition rate=14%	UCLA[c] loneliness scale [47]
van der Heide et al (2012) [48], The Netherlands	Low	Before and after study (with intervention group only, no control group)	Older home care	Baseline: average age 73.2 (SD 11.8), range 32-90; end of study: average age 73.1 (SD 11.2), range 38-90	Baseline: male=26 (30.2%), female=60 (69.8%), missing values=44; end of study: male=25 (29.4%), female=60 (70.6%), missing values=0	Not reported	Loneliness and safety issues	Convenience	130	130	85=intervention group at the end of study; no control group	45; attrition rate=34.6%	De Jong-Gierveld loneliness scale (score range: 0-11) [49]

Study, country	Quality of evidence[a] (reviewers' assessment)	Research design	Settings	Participants			Main health or medical conditions investigated	Sampling method	Sample size			Participant attrition	Research methods or data collection tools
				Age (years)	Gender	Ethnicity			Total	Intervention group	Control group		Loneliness scale used
Larsson et al (2016) [50], Sweden	High	Randomized, crossover trial	Living in ordinary housing without any home care services	Range: 61-89, mean 71.2; group 1 (intervention or control group): range 66-89, mean 73.4; group 2 (control or intervention group): range 61-76, mean 69.0	Male=6; female=24, (3 males and 12 females each in group 1 [intervention or control group] and group 2 [control or intervention group])	Not reported (probably all Swedes)	Loneliness	Randomized (after recruitment)	30	15 baseline, 14 follow-up	15 baseline, 14 follow-up	2 (1 participant each from intervention and control groups); attrition rate=6.7%	UCLA loneliness scale [51], the Swedish version (20 items, score range 20-80) [52]
Czaja et al (2018) [45], United States	High	Multisite randomized controlled trial	Living in independent housing in the community	Baseline: total sample mean 76.15 (SD 7.4), range: 65-98; intervention (PRISM[d] System) group: mean 76.9 (SD 7.3); control (Binder[e]) group: mean 75.3 (SD 7.4)	Baseline: female=78% (number not reported), male=22% (number not reported); PRISM or intervention group: female 79.3% (n=119); Binder (control) group: female 76.7% (n=115)	Baseline: White=54% and non-White=46%; PRISM or intervention group: non-White or Hispanic=8% (n=12); Binder group: non-White or Hispanic=10% (n=15)	Social isolation, social support, loneliness, and well-being	Randomized	300 (150 in each intervention [PRISM] group and control [Binder] group)	150 baseline; 134 follow-up	150 baseline; 118 follow-up	56 (45 at 6 months and 11 at 12-month follow-up); attrition rate=18.7%	UCLA loneliness scale (score range 20-80) [51]

Study, country	Quality of evidence[a] (reviewers' assessment)	Research design	Settings	Participants			Main health or medical conditions investigated	Sampling method	Sample size			Participant attrition	Research methods or data collection tools
				Age (years)	Gender	Ethnicity			Total	Intervention group	Control group		Loneliness scale used
Morton et al (2018) [53], United Kingdom	High	2 (condition: training, control)×2 (population: domiciliary, residential)×2 (time: baseline, follow-up) design	Receiving care in own home or supported housing in the community (*domiciliary care*) or residential care in care homes	Female: mean 80.71 (SD 8.77); male: data not reported	Follow-up: total=76; female=50, male=26	Not reported	Well-being and social support	Randomized	97 baseline; 76 follow-up	53 baseline; 44 follow-up	44 baseline; 32 follow-up	21 (9 experimental group; 12 control group); attrition rate=21.6%	UCLA loneliness scale (score range 20-80) [51]
Jarvis et al (2019) [54], South Africa	High	Randomized control trial study	Inner-city residential; NGO[f] care facilities for resource-restricted older people (aged ≥60 years)	Mean 74.93 (SD 6.41); range 61-87	Baseline: male=6 (18.8%), female=26 (81.2%)	Mostly Asian (of Indian origin), numbers not reported	Maladaptive cognitions and loneliness	Randomized	Baseline=32 (intervention group=15, control group=17), final=29 (intervention group=13, control group=16)	15 baseline; 13 follow-up	17 baseline; 16 follow-up	3 (2 intervention group, 1 control group); attrition rate=15.6%	De Jong-Gierveld loneliness scale (score range 0-11) [49]

[a] Quality of evidence grades: high (we are very confident that the true effect lies close to that of the estimate of the effect), moderate (we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different), low (our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect), and very low (we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect).

[b] NRCT: nonrandomized clinical trial.

[c] UCLA: University of California, Los Angeles.

[d] PRISM: Personal Reminder Information and Social Management.

[e] Binder refers to a group of participants who received a notebook with printed content similar to the Personal Reminder Information and Social Management System.

[f] NGO: nongovernmental organization.

Table 2. Interventions, outcomes, measurements, results, and conclusions of included studies.

Study	Interventions	Comparators	Intervention duration	Follow-up duration	Outcomes: loneliness scores by measurement stages, mean (SD)					Results or findings	Conclusion by the authors of the study
					Baseline	3 months	4 months	6 months	12 months		
Tsai et al (2010) [46]	Videoconferencing (using either MSN[a] messenger or Skype)	Regular care	3 months	3 months	Intervention group=50.58 (SD 11.16); control group=46.55 (SD 9.07)	Intervention group=47.33 (SD 13.50); control group=46.68 (SD 9.08)	Not measured	Not measured	Not measured	Loneliness: intervention group mean: baseline 50.58 (SD 11.16), 1 week 49.75 (SD 11.79), and 3 months 47.33 (SD 13.50); control group mean: baseline 46.55 (SD 9.07), 1 week 47.06 (SD 8.75), and 3 months 46.68 (SD 9.08); differences between groups were compared at 3 points (baseline, 1 week, and 3 months) using multiple linear regression of the generalized estimating equations. Unadjusted or fixed effect size of effectiveness of videoconferencing intervention (videoconference vs control): at 1 week was $\beta=-1.21$, SE 0.50, $\chi^2=5.9$, $P=.02$ and at 3 months $\beta=-2.84$, SE 1.28, $\chi^2=4.9$, $P=.03$	Videoconferencing alleviates depressive symptoms and loneliness in older residents in nursing homes
van der Heide et al (2012) [48]	CareTV including Caret duplex video or voice network	No control group and no comparator	12 months	12 months	Intervention group=5.97 (SD 2.77); no control group	Not measured	Not measured	Not measured	Intervention group=4.02 (SD 3.91); no control group	Group-level total loneliness: inclusion stage: mean 5.97 (SD 2.77), end of study: mean 4.02 (SD 3.91), $P=.001$; individual-level total loneliness: total loneliness decreased in 54 out of 85 participants (equally lonely 11, more lonely 20, and less lonely 54 individual participants)	CareTV intervention decreased the feeling of loneliness in the participants; however, participants were feeling moderate loneliness at the end of the study

Study	Interventions	Comparators	Intervention duration	Follow-up duration	Outcomes: loneliness scores by measurement stages, mean (SD)					Results or findings	Conclusion by the authors of the study
					Baseline	3 months	4 months	6 months	12 months		
Larsson et al (2016) [50]	SIBAs[b], that is, social activities via social websites	No comparator intervention reported	3 months	34 weeks (exposure for 3 months to each group)	Group 1 (I/C[c] group)=45.53 (SD 7.41); group 2 (C/I[d] group)=43.93 (SD 8.61)	Group 1 (I/C group)=42.43 (SD 7.44); group 2 (C/I group)=41.93 (SD 8.82)	Not measured	3 months after cross over: group 1 (I/C group, no intervention)=42.0 (SD 7.34); group 2 (C/I group, intervention introduced)=39.50 (SD 10.42)	Not measured	Percentage change between time 2 and time 1: group 1: mean score 0.07% (SD 0.07), *P*=.003; group 2: mean score: 0.05% (SD 0.09), *P*=.049; percentage change between time 3 and time 1: group 1: mean score 0.08% (SD 0.08); group 2: mean score 0.09% (SD 0.13); comparison of pre and postintervention scores: group 1, *P*=.003 and group 2, *P*=.049	SIBA interventions have the potential to reduce experiences of loneliness in socially vulnerable older adults.
Czaja et al (2018) [45]	PRISM[e] system	A notebook with printed content similar to that within the PRISM (intervention) group: included a Lenovo *Mini Desktop* PC with a keyboard, mouse (or trackball for those who were unable to control a mouse), a 19″ LCD[f] monitor, the PRISM software app, a printer, and internet	12 months	12 months	Intervention (PRISM) group=39.8 (SD 9.7); control (Binder[g]) group=40.2 (SD 10.3)	Not measured	Not measured	Intervention (PRISM) group=37.8 (SD 9.54); control (Binder) group=40 (SD 10.62)	Intervention (PRISM) group=36.9 (SD 9.16); control (Binder) group=38.43 (SD 9.37)	Baseline: loneliness PRISM group: mean score 39.8 (SD 9.7); Binder group: mean score 40.2 (SD 10.3), follow-up at 6 months: PRISM group 37.8, Binder group 39.6; follow-up at 12 months: PRISM group 36.9, Binder group 38.3	Technology-based apps such as the PRISM system may enhance social connectivity and reduce loneliness among older adults.

Study	Interventions	Comparators	Intervention duration	Follow-up duration	Outcomes: loneliness scores by measurement stages, mean (SD)					Results or findings	Conclusion by the authors of the study
					Baseline	3 months	4 months	6 months	12 months		
Morton et al (2018) [53]	EasyPC—a customized computer platform with a simplified touchscreen interface	Care as usual plus regular carer visits	3 months	4 months	Intervention (training) group (total of residential and domiciliary groups)=1.92 (SE 0.10, SD 0.73); control group (total of residential and domiciliary groups)=2.08 (SE 0.12, SD 0.80)	Not measured	Intervention (training) group (total of residential and domiciliary groups)=1.86 (SE 0.10, SD 0.66); control group (total of residential and domiciliary groups)=2.12 (SE 0.11, SD 0.62)	Not measured	Not measured	Loneliness scores mean: intervention (training) group: residential group: time 1=1.95 (SE 0.16), time 2=1.92 (SE 0.16), domiciliary group: time 1=1.89 (SE 0.13), time 2=1.79 (SE 0.13), total time 1=1.92 (SE 0.10), time 2=1.86 (SE 0.10); control group: residential group: time 1=2.13 (SE 0.18), time 2=2.20 (SE 0.17), domiciliary group: time 1=2.02 (SE 0.16), time 2=2.05 (SE 0.15), total time 1=2.08 (SE 0.12) and time 2=2.12 (SE 0.11)	Internet access and training can support the self and social connectedness of vulnerable older adults and contribute positively to well-being.
Jarvis et al (2019) [54]	Living In Network-Connected Communities WhatsApp group for low-intensity cognitive behavioral therapy	Usual care, a separate WhatsApp group (Living In Network-Connected Communities 2)	3 months	4 months	Not measured	Intervention group=2.31 (SD 1.49); control group=2.47 (SD 2.1)	Intervention group=1.38 (SD 1.33); control group=4.0 (SD 1.32)	Not measured	Not measured	Loneliness levels: total=baseline−intervention on time 1−intervention on time 2.; χ^2=14.6; *P*=.001	Low-intensity cognitive behavioral therapy mobile health supported by the social networking platform of WhatsApp (Living In Network-Connected Communities) showed significant improvements in loneliness and maladaptive cognitions.

[a] MSN: Microsoft Network.

[b] SIBA: social internet-based activity.

[c] I/C: intervention/control.

[d] C/I: control/intervention.

[e] PRISM: Personal Reminder Information and Social Management.

[f] LCD: liquid-crystal display.

[g] Binder refers to a group of participants who received a notebook with printed content similar to the Personal Reminder Information and Social Management System.

SGSS and DN independently extracted data from all included studies (n=6) using the data extraction template (Tables 1 and 2) and resolved discrepancies in the extracted data with discussion and agreement. Data extraction forms were compared and contrasted, thereby avoiding bias and reducing errors in the data extraction process [55]. We extracted aggregated data at the study level as much as possible with respect to the intervention, which is imperative for the reproducibility of effective interventions [56,57]. Following suggestions for reporting data once from studies with duplicate and multiple publications [55], we extracted and reported data only once [50] from a research study with multiple publications [50,58].

Data Synthesis and Reporting

We report both a narrative synthesis (narrative summary) and a statistical (quantitative) synthesis (meta-analysis) of our review, as suggested for reporting of a systematic review on effectiveness [59]. In the narrative synthesis, we have included all 6 studies and reported their characteristics, including the study design, settings, sample sizes, data collection methods, participants, interventions, comparators, outcome measurements, and study conclusions.

In the meta-analysis, we have included 5 studies and pooled extracted data on loneliness measured by continuous loneliness scales, that is, the University of California, Los Angeles (UCLA) loneliness scale [47,51] in 4 studies and the De Jong-Gierveld loneliness scale [49] in 2 studies (Table 1). Loneliness scores at baseline and follow-up were reported as the mean values and SDs in 5 studies, whereas 1 study reported mean scores with SEs. For the latter study, we calculated SDs from SEs using a formula suggested in the Cochrane guidelines [60].

In meta-analysis, the standardized mean difference (SMD) as a summary statistic for reporting continuous data has been suggested for studies that assess the same outcome but use different scales to measure the outcome [60]. In RevMan (The Cochrane Collaboration), the SMD is the effect size known as Hedges (adjusted) g, which is akin to Cohen *d* and includes an adjustment for small sample size bias [60]. More importantly, the generalizability of the SMD statistic is more than the mean difference statistic in a meta-analysis [61].

In our review, the main outcome, that is, loneliness, was measured using different loneliness scales, which included the UCLA loneliness scale (score range 20-80) [47,51] and the De Jong-Gierveld loneliness scale (score range 0-11) [49]. Although these 2 loneliness scales have commonalities such as self-reporting measures and focus on the functional dimension of social relationship and the degree of subjectivity covering perceived availability, adequacy, and emotions or feelings, they differ from each other in other aspects, such as the content and formulation of items or questions included in the scales [62]. In addition, the 2 measures have different number of items or questions, rating options, scoring methods, total scores, and scale versions (for details, refer to the studies by Russell [51] and Russell et al [47] for the UCLA loneliness scale and the studies by De Jong-Gierveld and Tilburg [49] and De Jong-Gierveld and Kamphuls [63] for the De Jong-Gierveld loneliness scale).

The Cochrane guidelines for systematic reviews and meta-analysis [60] suggest that different study designs should not be combined in a meta-analysis because it can increase heterogeneity, and studies with repeated measurements at different follow-up periods cannot be combined without a unit of analysis error.

We extracted data from 6 studies, which included 5 clinical trials [45,46,50,53,54] and 1 pre-post study [48]. Therefore, we included similar study designs, that is, clinical trials in the meta-analysis, and conducted separate meta-analyses based on the same follow-up measurement periods in the clinical trials. Therefore, we performed a separate meta-analysis for each follow-up, that is, measurements at 3, 4, and 6 months after the intervention. In addition, we ran meta-analyses when there were at least two or more studies for the same outcome or the same follow-up period [64]. Therefore, we did not conduct a meta-analysis for follow-up measurements at 12 months reported in 2 studies because they involved different study designs, that is, a randomized controlled trial (RCT) with intervention and control groups [45] and a pre- and postintervention study with only intervention group [48]. This was done to avoid an increase in the heterogeneity [60] and overestimation of the effect of intervention in the absence of a control group [65] in the pre- and postintervention study [48]. We did not perform a meta-analysis for the pre-post study [48] because meta-analysis cannot be performed with only 1 study [60].

We calculated the SMDs from the extracted data, that is, loneliness mean scores with SD and sample sizes in the intervention and control groups at follow-up measurements at 3 months and beyond. For conducting meta-analysis, we used the Cochrane Review Manager (RevMan) software, version 5.3.5 [66]. In the meta-analysis, we used the random effects model as the statistical model because we hypothesized that the true effect sizes between studies would vary [67,68] due to differences in the methodological and clinical characteristics between studies [69], such as differences in the sample sizes, participant numbers and characteristics, intervention types and durations, and follow-up measurement times. We did not conduct sensitivity analyses because of the small number of studies in the meta-analyses at each follow-up point [64].

Assessment of Research Quality, Bias, and Heterogeneity

We assessed the quality of research by applying the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach [70].

We assessed the risk of bias by focusing on 5 domains: the evaluation of sequence generation, allocation concealment, blinding (outcome assessors), incomplete data, selective outcome reporting, and assessing other biases using the Cochrane guidelines [60]. In a meta-analysis, publication bias can be assessed with a graphical method using funnel plots [60,71] and statistical methods such as the Egger test [60]; however, both methods require at least 10 studies in the meta-analysis [60]. When the number of studies is small, the

Egger test has low power and fails to differentiate chance from real asymmetry [60]. Similarly, assessing publication bias using funnel plots with fewer studies would be of very limited usefulness because it would be difficult to spot the publication bias. As we had a maximum of 3 studies in a meta-analysis, we could not check the publication bias with either method.

We checked heterogeneity, that is, variation in study outcomes or intervention effect sizes between studies, by the Cochran Q test with a significance level of $\rho<0.10$ [72,73] because of the low power of the test in a meta-analysis with very few studies or studies with small sample sizes [74]. We calculated I^2 statistics to determine the magnitude of heterogeneity (ie, the proportion of variance in the true effect sizes) between studies [28]. We considered I^2 values of 25%, 50%, and 75% as low, moderate, and high heterogeneity between studies, respectively [75].

Summary Measures

We report the findings of meta-analyses using SMDs with 95% CIs as a statistical summary, with the forest plots [60].

Results

Narrative Synthesis

Findings about the characteristics of the studies, including the study designs, settings, participants, interventions, comparators, sample sizes, participant attrition, and data collection methods or tools used, are presented in Table 1. The interventions, comparators, follow-up durations, outcomes or measurement scores, results, and conclusions of the included studies are given in Table 2.

Study Selection

Searches of PubMed, MEDLINE, CINAHL, Embase, and Web of Science generated a total of 4939 articles (Figure 1), of which 6 studies met the predefined eligibility criteria. All 6 studies were included in the narrative synthesis, 5 clinical trials with the intervention and control groups were included in the meta-analysis, and only 1 study with a pre-post design involving only the intervention group was excluded from the meta-analysis.

Study Participants

The total number of participants enrolled in all 6 included studies was 646 (mean 108, SD 102; median 77, IQR 32-130). Studies varied in total sample sizes (mean 108, SD 102; range 30-300), and the sample sizes of the intervention and control groups also varied at both the baseline and follow-up measurements across the studies (Table 1). The attrition rate also varied between studies (range 7%-35%; mean 19%, SD 10%).

Participants’ average age was between 73 and 78 years (SD 6-11). Total enrolled participants included 66.1% (427/646) women and 23.8% (154/646) men, whereas for 10.1% (65/646) of participants, no information about their gender was available. Studies varied in the proportion of male and female participants (female: mean 66%, SD 16%; range 46%-81%; male: mean 25%, SD 9%; range 19%-42%). Only 2 studies reported on participants’ ethnicity—White (54%) and non-Whites (46%) in the US study [45] and mostly Asian Indians (no numbers reported) in the South African study [54].

Study Characteristics

A total of 4 studies were RCTs [45,50,53,54], 1 study was a nonrandomized clinical trial [46], and 1 was a pre- and posttest (before and after) study with intervention group only (no control group) [48] (Table 1).

Study Settings

A total of 4 studies were conducted in developed countries, namely, the Netherlands [48], the United Kingdom [53], the United States [45]**,** and Sweden [50]. Two studies were undertaken in developing countries, namely, Taiwan [46] and South Africa [54].

The settings included living in independent housing in the community [45]; living in ordinary housing without any home care services [50]; receiving care in their own home or supported housing in the community (*domiciliary care*), or receiving care in residential care homes [53], residential care facilities for older people [54], nursing homes [46], and older home care [48].

Participants were selected by random sampling in 66.7% (4/6) of studies [45,50,53,54], whereas the other 33.3% (2/6) studies used purposive [46] and convenience [48] sampling each.

Digital Technology Interventions

DTIs included social internet-based activities, that is, social activities via social websites [50], videoconferencing [46], customized computer platforms with simplified touch-screen interfaces [53], personal reminder information and social management systems [45], WhatsApp groups [54], and video or voice networks [48].

Duration of the Intervention and Measurement of the Main Outcome Measure

The duration of the intervention was 3 months in 4 studies [46,50,53,54] and 12 months in 2 studies [45,48]. The main outcome measure, that is, loneliness, was measured at the baseline and multiple follow-up times, which included 3 months in 3 studies [46,50,54], 4 months in 2 studies [53,54], 6 months in 2 studies [45,50], and 12 months in 2 studies [45,48].

The loneliness measurement tools used were the UCLA loneliness scale [47,51], which was applied in 4 studies [45,46,50,53], and the De Jong-Gierveld loneliness scale [49,76], which was used in 2 studies [48,54]. Table 2 presents loneliness scores measured in the intervention and control groups, if any, at baseline and follow-ups.

Narrative synthesis showed that there was a reduction in loneliness in the intervention groups at the follow-ups compared with baseline (Table 2). A statistical summary of the loneliness measurements in the intervention and control groups at the follow-ups is reported in the *Meta-analysis* section.

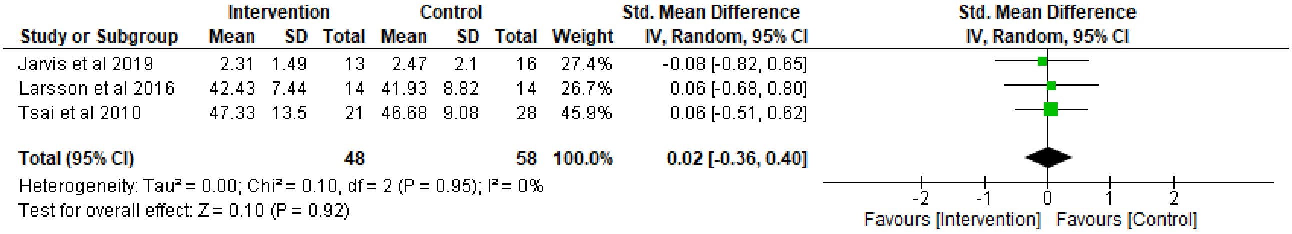
Meta-analysis

We conducted 3 meta-analyses, 1 each for follow-up measurements at 3, 4, and 6 months, involving 3, 2, and 2 studies, respectively.

Meta-analysis for Follow-up at 3 Months

Three studies [46,50,54] involving 106 participants with follow-up measurements at 3 months were entered into a meta-analysis, which showed a very small reduction in loneliness in favor of the control (SMD 0.02; 95% CI −0.36 to 0.40), but it was not statistically significant (Z=0.10; *P*=.92). The heterogeneity between studies was not statistically significant (τ^2=0.00; χ^2_2=0.1; *P*=.95; I^2=0%; Figure 2).

Figure 2. Forest plot of standardized mean differences for loneliness at the 3-month follow-up (digital technology intervention vs control).



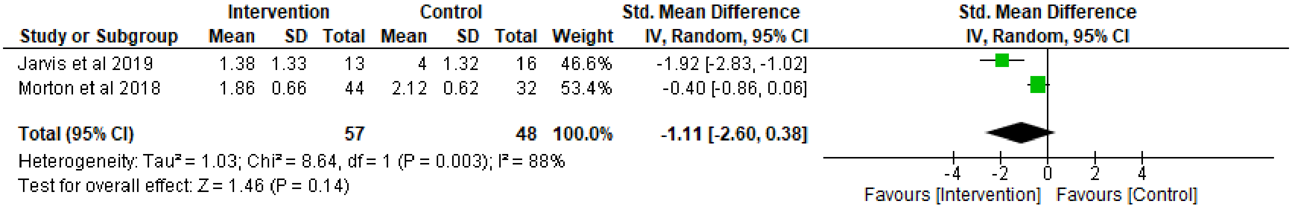
Study or Subgroup	Intervention Mean	Intervention SD	Intervention Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Jarvis et al 2019	2.31	1.49	13	2.47	2.1	16	27.4%	-0.08 [-0.82, 0.65]
Larsson et al 2016	42.43	7.44	14	41.93	8.82	14	26.7%	0.06 [-0.68, 0.80]
Tsai et al 2010	47.33	13.5	21	46.68	9.08	28	45.9%	0.06 [-0.51, 0.62]
Total (95% CI)			**48**			**58**	**100.0%**	**0.02 [-0.36, 0.40]**

Heterogeneity: Tau² = 0.00; Chi² = 0.10, df = 2 (P = 0.95); I² = 0%
Test for overall effect: Z = 0.10 (P = 0.92)

Meta-analysis for Follow-up at 4 Months

Two studies [53,54] involving 105 participants with 4 month follow-up were entered into a meta-analysis, which revealed a large reduction in loneliness in favor of the intervention (SMD −1.11; 95% CI −2.60 to 0.38), but it was not statistically significant (Z=1.46; *P*=.14). There was a statistically significant high heterogeneity between studies (τ^2=1.03; χ^2_1=8.8; *P*=.003; I^2=88%; Figure 3).

Figure 3. Forest plots of standardized mean differences for loneliness at the 4-month follow-up (digital technology intervention vs control).



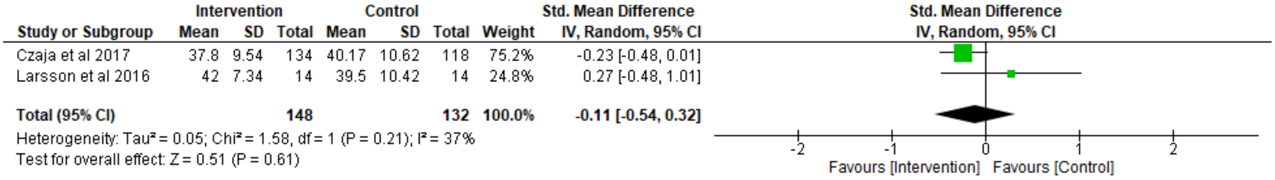
Study or Subgroup	Intervention Mean	Intervention SD	Intervention Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Jarvis et al 2019	1.38	1.33	13	4	1.32	16	46.6%	-1.92 [-2.83, -1.02]
Morton et al 2018	1.86	0.66	44	2.12	0.62	32	53.4%	-0.40 [-0.86, 0.06]
Total (95% CI)			**57**			**48**	**100.0%**	**-1.11 [-2.60, 0.38]**

Heterogeneity: Tau² = 1.03; Chi² = 8.64, df = 1 (P = 0.003); I² = 88%
Test for overall effect: Z = 1.46 (P = 0.14)

Meta-analysis for Follow-up at 6 Months

A meta-analysis involving 2 studies [45,50] with 280 participants with 6 month follow-up showed a very small reduction in loneliness in favor of the intervention (SMD −0.11; 95% CI −0.54 to 0.32), but it was not statistically significant (Z=0.51; *P*=.61). There was moderate heterogeneity between studies, but it was not statistically significant (τ^2=0.05; χ^2_1=1.6; *P*=.21; I^2=37%; Figure 4).

Figure 4. Forest plots of standardized mean differences for loneliness at the 6-month follow-up (digital technology intervention vs control).



Study or Subgroup	Intervention Mean	Intervention SD	Intervention Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Czaja et al 2017	37.8	9.54	134	40.17	10.62	118	75.2%	-0.23 [-0.48, 0.01]
Larsson et al 2016	42	7.34	14	39.5	10.42	14	24.8%	0.27 [-0.48, 1.01]
Total (95% CI)			**148**			**132**	**100.0%**	**-0.11 [-0.54, 0.32]**

Heterogeneity: Tau² = 0.05; Chi² = 1.58, df = 1 (P = 0.21); I² = 37%
Test for overall effect: Z = 0.51 (P = 0.61)

Risk of Bias

The risk of bias assessment, that is, the risk of bias graph and risk of bias summary are presented in Figure 5 and Figure 6, respectively. A high risk of bias was noted in the attrition bias and other biases; an unclear risk of bias was detected in the blinding of outcome assessment, allocation concealment, and blinding of participants and personnel; and a low risk of bias was observed, especially, in the random sequence generation and selective reporting (Figure 6). In addition, most studies reported only within-group changes and not between-group comparisons of change, which may suggest a weak quality of the reporting of results and the analysis in these studies (Table 2).

Figure 5. Risk of bias summary. Review authors' judgments about risk of bias in included studies: Czaja et al, 2017 [45], Tsai et al 2010 [46], Larsson et al, 2016 [50], Morton et al, 2018 [53], and Jarvis et al, 2019 [54].

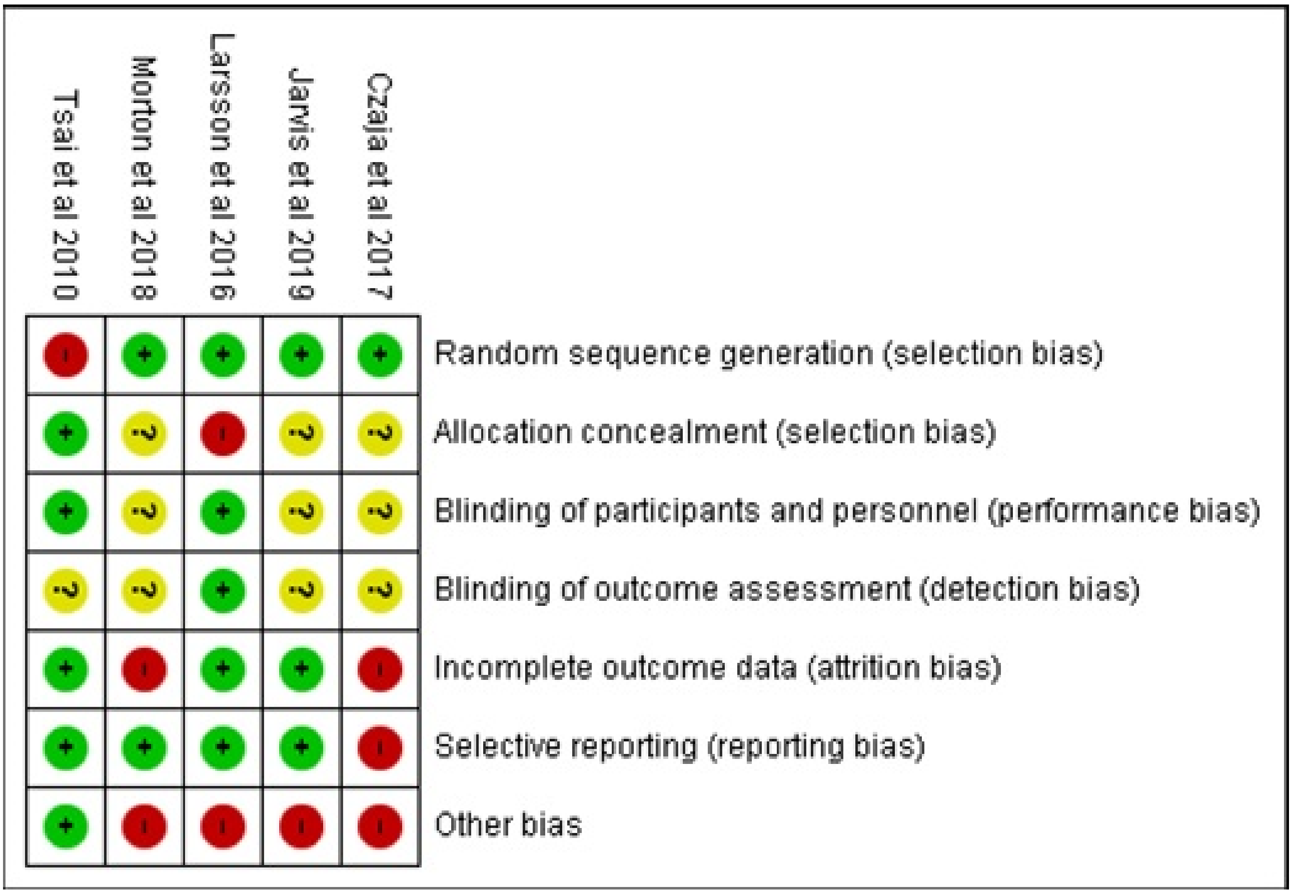
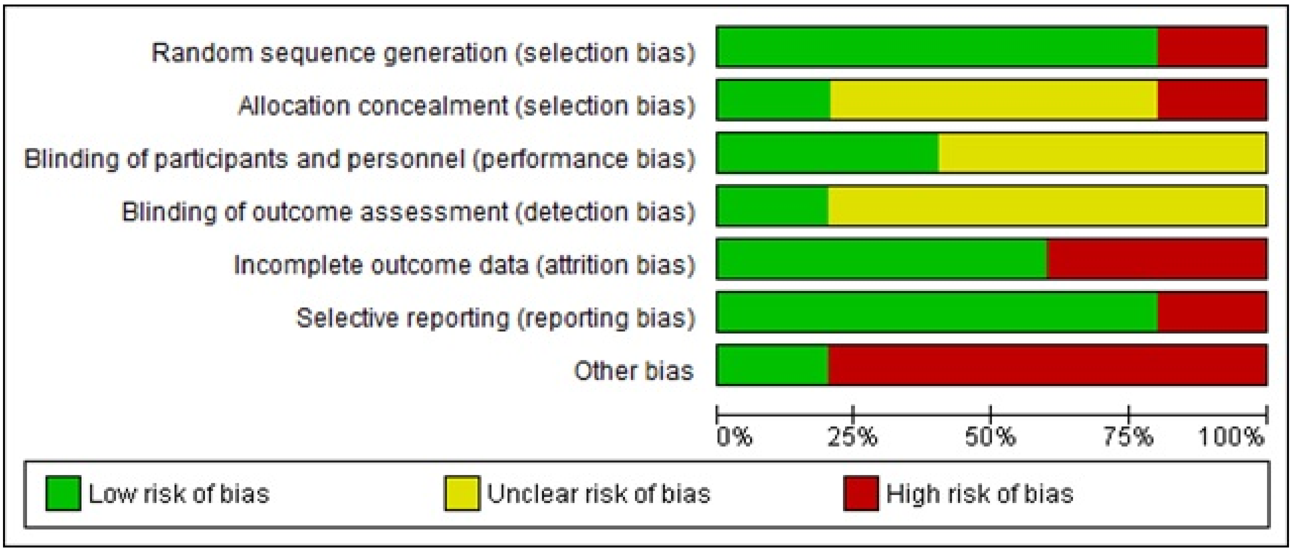


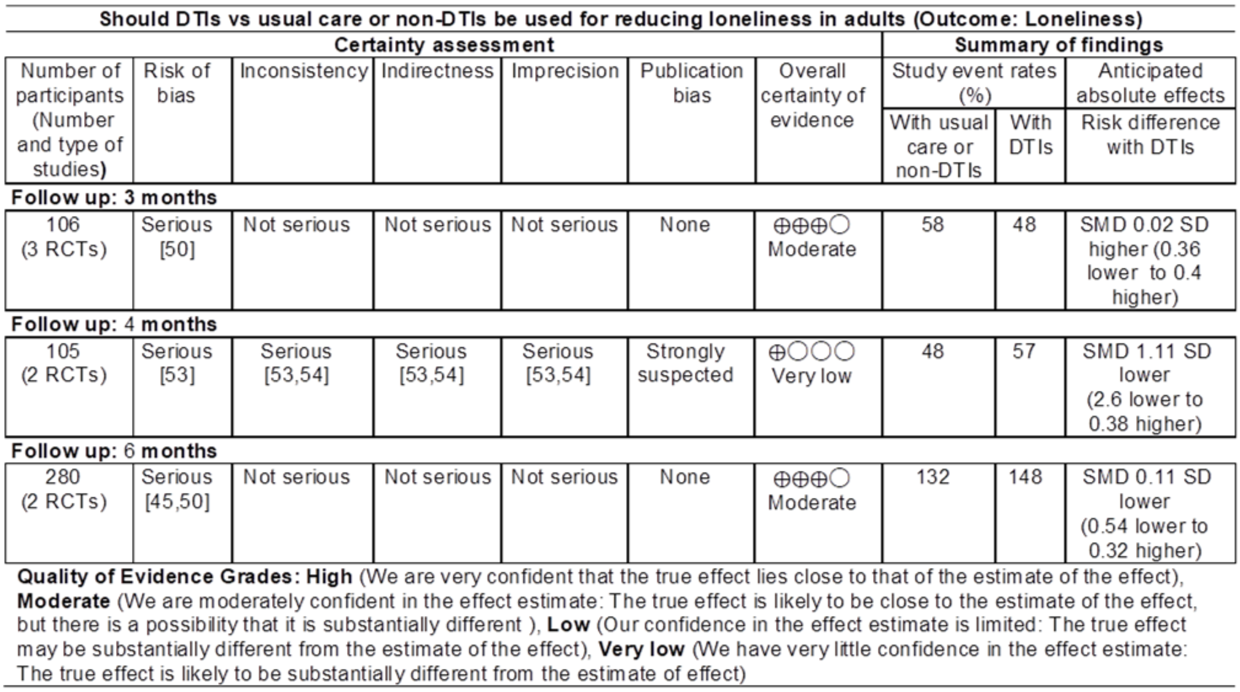
Figure 6. Risk of bias graph. Review authors' judgments about each risk of bias item are presented as percentages across all included studies.



Quality of Evidence

The quality of evidence was moderate, very low, and moderate in meta-analyses involving 3 [46,50,54], 2 [53,54], and 2 studies [45,50], with follow-up at 3, 4, and 6 months, respectively (Figure 7).

Figure 7. GRADE (Grading of Recommendations Assessment, Development and Evaluation) quality of evidence summary. DTI: digital technology intervention; RCT: randomized controlled trial; SMD: standardized mean difference.



Should DTIs vs usual care or non-DTIs be used for reducing loneliness in adults (Outcome: Loneliness)

Certainty assessment							**Summary of findings**		
Number of participants (Number and type of studies)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall certainty of evidence	Study event rates (%)		Anticipated absolute effects
							With usual care or non-DTIs	With DTIs	Risk difference with DTIs
Follow up: 3 months									
106 (3 RCTs)	Serious [50]	Not serious	Not serious	Not serious	None	⊕⊕⊕○ Moderate	58	48	SMD 0.02 SD higher (0.36 lower to 0.4 higher)
Follow up: 4 months									
105 (2 RCTs)	Serious [53]	Serious [53,54]	Serious [53,54]	Serious [53,54]	Strongly suspected	⊕○○○ Very low	48	57	SMD 1.11 SD lower (2.6 lower to 0.38 higher)
Follow up: 6 months									
280 (2 RCTs)	Serious [45,50]	Not serious	Not serious	Not serious	None	⊕⊕⊕○ Moderate	132	148	SMD 0.11 SD lower (0.54 lower to 0.32 higher)

Quality of Evidence Grades: High (We are very confident that the true effect lies close to that of the estimate of the effect), **Moderate** (We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different), **Low** (Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect), **Very low** (We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect)

Discussion

Principal Findings

To determine whether DTIs are effective in reducing loneliness in adults, we appraised peer-reviewed empirical research involving the application of DTIs in adults with loneliness. Our systematic review provides a narrative summary (qualitative synthesis) as well as a meta-analysis (statistical synthesis) of the findings. The narrative summary of 6 studies included in our review showed a reduction in loneliness in the intervention groups at follow-up compared with baseline (Table 2). However, our meta-analysis of 5 clinical trials with follow-up measurements at 3, 4, and 6 months showed no statistically significant pooled effect estimates as SMDs, the preferred method for summarizing effects on continuous outcomes such as loneliness. Although not statistically significant, the summary effect size at the 4-month follow-up (Figure 3) was better than the effect size at the 3-month follow-up (Figure 2) and the 6-month follow-up (Figure 4).

Our meta-analysis also revealed that CIs of the summary effects of 2 studies, that is, the studies by Larsson et al [50] and Tsai et al [46], were very wide, and the SMDs from these studies were more in favor of the control group than the intervention group (Figure 2). Thus, the wide range of CIs of the summary effects in these studies leave room for uncertainty about the beneficial effect of DTIs on measures of loneliness.

Overall, the findings of our meta-analysis showed no evidence supporting the effectiveness of DTIs in reducing loneliness in older adults.

Summary of Evidence

The quality of evidence of the included studies was very low to moderate (Figure 7), and there was a high heterogeneity between studies [53,54] (Figure 2). All the included studies had a high proportion of female participants. Most notably, the total number of participants was low, especially in 2 studies [50,54], and the sample sizes were reduced further due to a high attrition rate in some studies [45,53]. The types and methods of DTIs varied between studies (Table 2), which were conducted in diverse settings (Table 1). Studies were conducted in 6 different countries: the United States, the United Kingdom, Sweden, the Netherlands (these four countries have individualistic cultures), Taiwan, and South Africa (these two countries have collectivist cultures; Table 1). In addition, despite our inclusion criteria of age 18 years and above, the selected studies more commonly involved older people with an average age of 70 years and above.

Loneliness is influenced by culture [77-79], gender [78], and age [5,78], and these factors could have contributed to the pooled estimates being not statistically significant in our meta-analysis. In addition, differences in participants, especially in terms of age, gender, and culture as well as varied types of DTIs, could have contributed to the heterogeneity observed, especially in the meta-analysis with the 4 month follow-up involving 2 studies [53,54], which differed from each other on different parameters, especially the study designs, settings, participants, interventions, and loneliness measurement scales used (Tables 1 and 2).

There are limited published meta-analyses on technological interventions for tackling loneliness, and a few existing studies have covered literature published up to 2009 [28] and 2011 [30]. Our review and meta-analysis included the latest evidence published between January 1, 2010, and July 31, 2019. We did not replicate the findings of earlier meta-analyses that reported evidence suggesting that technological interventions resulted in decreased loneliness [28,30]. For example, a meta-analysis by Choi et al [30] reported statistically significant evidence suggesting that the internet and computers reduce loneliness. However, they [30] focused on older adults with depression and included the internet and computers only as technological interventions, whereas we included different types of DTIs, and our population of interest was adults of all age groups (≥18 years). In addition, the meta-analysis by Choi et al [30] included studies (n=5) with different follow-up periods (3-6 months), but they did not report which follow-up measurements were included in their meta-analysis. In our meta-analysis, we conducted separate meta-analyses for measurements at different follow-up periods, that is, 3, 4, and 6 months, as suggested by the Cochrane guidelines [60].

A meta-analysis by Masi et al [28] also reported that technological interventions reduce loneliness, which was more in pre-post studies and nonrandomized studies than in RCTs. However, they included studies with technology and nontechnology-based interventions [28], whereas we focused on studies with DTIs only. In addition, Masi et al [28] did not report how they analyzed measurements at different follow-up periods, whereas we did not combine measurements at different follow-up times, as suggested by the Cochrane guidelines [60]. Nonetheless, Masi et al [28] concluded that technology is yet to be capitalized for loneliness.

Interestingly, our findings provide new insights about DTIs and loneliness. Our meta-analysis showed no statistically significant reduction in loneliness in the intervention groups compared with the control groups at the 3 -, 4-, and 6-month follow-ups. Thus, our findings show no evidence supporting the effectiveness of DTIs in reducing loneliness in older adults, which goes beyond the findings of a recent Cochrane review that reported no evidence of video calls being effective in reducing loneliness in older adults [73].

In addition, our findings refute and contradict a commonly held view that digital technology can solve the problem of loneliness, especially in older people. Nonetheless, digital technologies provide tools and means that facilitate social connection [80], which may help in reducing loneliness for a limited period because the effects of DTIs are short-lived [81]. This may be because digital technologies do not provide real human interaction [80] and cannot replace human contact [45]; thus, they do not reduce social disconnectedness in real life [82] on a long-term basis.

Nonetheless, a review has reported that some nontechnological interventions are effective in reducing loneliness in older people [83], but these interventions require a meta-analytic evaluation. In addition, a recent meta-analysis [84] reported moderate evidence of the effectiveness of a range of social, emotional, and psychological interventions, delivered through technological and nontechnological means, in reducing loneliness in young people aged 3-25 years; however, the analyzed studies had limitations. Therefore, further research is required.

Limitations

Our study had some limitations: the inclusion of only 6 studies with heterogeneous sets of results and the minimum intervention duration of 3 months, which could have resulted in the inclusion of a small number of studies and possible exclusion of potential studies that would have provided useful evidence.

In addition, we could not conduct subgroup and meta-regression analyses due to the very limited number of studies (n=5) in the meta-analysis and lack of data on loneliness by participants' demographic characteristics. In addition, our study might be narrow because we excluded some studies [44,85-91], which met the technology criterion such as the use of robots, sensors, digital speakers, and apps but did not meet other selection criteria. Thus, our study may be limited to studies about social interactions and connectedness using digital technology tools.

Moreover, another limitation of our review could be the use of a meta-analysis based only on follow-up data. For example, a study by Tsai et al [46] in the 3-month follow-up meta-analysis had an SMD of 0.06 with a 95% CI of −0.8 to 0.65 (P=.03; Figure 2), which may suggest that these studies may have had higher power to show a difference compared with baseline loneliness.

As recommendations for future research, we suggest that researchers involved in trials agree on a common measure of loneliness and consider reporting of results in a standardized way, which will allow pooling of baseline-adjusted estimates of the treatment effect rather than differences in follow-up means.

Conclusions

Our meta-analysis showed no evidence supporting the effectiveness of DTIs in reducing loneliness in older adults. Therefore, there is a need for further research involving RCTs [50] with larger sample sizes and longer duration of interventions and follow-up measurement periods. Future research may apply inclusive research designs using a combination of digital apps, including robots, sensors, and social connecting apps, by involving adults aged 45-65 years, as this segment of the population is more likely to be more technology savvy and digital interventions might be more effective in this age group. Future research might also target ethnic minority communities and specific groups such as lesbian, gay, bisexual, and transgender people where loneliness is common [8,92].

Acknowledgments

The authors acknowledge support from Liz Callow of Bodleian Health Care Libraries, University of Oxford, for running literature searches. The authors are grateful to Dr Alexandra Farrow (Brunel University London) for checking the first version of the manuscript. This research was funded by the National Institute for Health Research Oxford Biomedical Research Centre (Research Grant IS-BRC-1215-20008). The views expressed are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature searches.
[DOCX File , 37 KB - jmir_v23i6e24712_app1.docx]

References

1. Lindsay EK, Young S, Brown KW, Smyth JM, Creswell JD. Mindfulness training reduces loneliness and increases social contact in a randomized controlled trial. Proc Natl Acad Sci U S A 2019 Feb 26;116(9):3488-3493 [FREE Full text] [doi: 10.1073/pnas.1813588116] [Medline: 30808743]
2. Loneliness is pervasive and rising, particularly among the young - daily chart. The Economist. 2018. URL: https://www.economist.com/graphic-detail/2018/08/31/loneliness-is-pervasive-and-rising-particularly-among-the-young [accessed 2019-05-03]
3. Mance P. Is Australian experiencing an epidemic of loneliness? Relationships Australia. Canberra, Australia: Relationships Australia National;; 2018 Sep. URL: https://www.relationships.org.au/pdfs/Anepidemicofloneliness20012017.pdf
4. van den Broek T. Gender differences in the correlates of loneliness among Japanese persons aged 50-70. Australas J Ageing 2017 Sep;36(3):234-237. [doi: 10.1111/ajag.12448] [Medline: 28685951]
5. Victor CR, Bowling A. A longitudinal analysis of loneliness among older people in Great Britain. J Psychol 2012;146(3):313-331. [doi: 10.1080/00223980.2011.609572] [Medline: 22574423]
6. Murthy V. Work and the loneliness epidemic. Harv Bus Rev. 2017. URL: https://hbr.org/cover-story/2017/09/work-and-the-loneliness-epidemic [accessed 2020-10-01]
7. Cohen-Mansfield J, Perach R. Interventions for alleviating loneliness among older persons: a critical review. Am J Health Promot 2015;29(3):109-125. [doi: 10.4278/ajhp.130418-LIT-182] [Medline: 24575725]
8. Jeste DV, Lee EE, Cacioppo S. Battling the modern behavioral epidemic of loneliness: suggestions for research and interventions. JAMA Psychiatry 2020 Jun 01;77(6):553-554. [doi: 10.1001/jamapsychiatry.2020.0027] [Medline: 32129811]
9. Gao Q, Prina M, Prince M, Acosta D, Sosa AL, Guerra M, et al. Loneliness among older adults in Latin America, China, and India: prevalence, correlates and association with mortality. Int J Public Health 2021 Mar 31;66:604449. [doi: 10.3389/ijph.2021.604449]
10. d’Hombres B, Schnepf S, Barjakovà M, Mendonça F. Loneliness – an unequally shared burden in Europe. European Commission. 2018. URL: https://ec.europa.eu/jrc/sites/jrcsh/files/fairness_pb2018_loneliness_jrc_i1.pdf [accessed 2020-09-15]
11. Galea S, Merchant RM, Lurie N. The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. JAMA Intern Med 2020 Jun 1;180(6):817-818. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]
12. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
13. Shah SG, Nogueras D, van Woerden HC, Kiparoglou V. The COVID-19 pandemic: a pandemic of lockdown loneliness and the role of digital technology. J Med Internet Res 2020 Nov 05;22(11):e22287 [FREE Full text] [doi: 10.2196/22287] [Medline: 33108313]
14. Okwaraji FE, Obiechina KI, Onyebueke GC, Udegbunam ON, Nnadum GS. Loneliness, life satisfaction and psychological distress among out-of-school adolescents in a Nigerian urban city. Psychol Health Med 2018 Oct;23(9):1106-1112. [doi: 10.1080/13548506.2018.1476726] [Medline: 29792068]
15. A Connected Scotland: our strategy for tackling social isolation and loneliness and building stronger social connections. The Scottish Government. 2018. URL: https://www.gov.scot/publications/connected-scotland-strategy-tackling-social-isolation-loneliness-building-stronger-social-connections/pages/6/ [accessed 2019-05-03]
16. Qualter P. Loneliness in children and adolescents: what do schools and teachers need to know and how can they help? Pastoral Care in Ed 2003 Jun;21(2):10-18. [doi: 10.1111/1468-0122.00257]
17. Cohen-Mansfield J, Hazan H, Lerman Y, Shalom V. Correlates and predictors of loneliness in older-adults: a review of quantitative results informed by qualitative insights. Int Psychogeriatr 2016 Apr;28(4):557-576. [doi: 10.1017/S1041610215001532] [Medline: 26424033]
18. Deckx L, van den Akker M, Buntinx F. Risk factors for loneliness in patients with cancer: a systematic literature review and meta-analysis. Eur J Oncol Nurs 2014 Oct;18(5):466-477. [doi: 10.1016/j.ejon.2014.05.002] [Medline: 24993076]
19. Chen Y, Hicks A, While AE. Loneliness and social support of older people in China: a systematic literature review. Health Soc Care Community 2014 Mar;22(2):113-123. [doi: 10.1111/hsc.12051] [Medline: 23714357]

20. Rico-Uribe LA, Caballero FF, Martín-María N, Cabello M, Ayuso-Mateos JL, Miret M. Association of loneliness with all-cause mortality: a meta-analysis. PLoS One 2018 Jan 4;13(1):e0190033 [FREE Full text] [doi: 10.1371/journal.pone.0190033] [Medline: 29300743]
21. Holt-Lunstad J. The potential public health relevance of social isolation and loneliness: prevalence, epidemiology, and risk factors. Public Policy Aging Rep 2017;27(4):127-130. [doi: 10.1093/ppar/prx030]
22. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
23. Deckx L, van den Akker M, Buntinx F, van Driel M. A systematic literature review on the association between loneliness and coping strategies. Psychol Health Med 2018 Sep;23(8):899-916. [doi: 10.1080/13548506.2018.1446096] [Medline: 29533084]
24. Lara E, Martín-María N, De la Torre-Luque A, Koyanagi A, Vancampfort D, Izquierdo A, et al. Does loneliness contribute to mild cognitive impairment and dementia? A systematic review and meta-analysis of longitudinal studies. Ageing Res Rev 2019 Jul;52:7-16. [doi: 10.1016/j.arr.2019.03.002] [Medline: 30914351]
25. Peytrignet S, Garforth-Bles S, Keohane K. Loneliness monetisation report : analysis for the Department for Digital, Culture, Media & Sport. 2020. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/910789/Loneliness_monetisation_report.pdf [accessed 2021-05-13]
26. Jeffrey K, Abdallah S, Michaelson J. The cost of loneliness to UK employers. New Economics Foundation. 2017. URL: https://neweconomics.org/2017/02/cost-loneliness-uk-employers/2017 [accessed 2020-05-20]
27. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. Proc Natl Acad Sci U S A 2013 Apr 9;110(15):5797-5801 [FREE Full text] [doi: 10.1073/pnas.1219686110] [Medline: 23530191]
28. Masi CM, Chen H, Hawkley LC, Cacioppo JT. A meta-analysis of interventions to reduce loneliness. Pers Soc Psychol Rev 2011 Aug;15(3):219-266 [FREE Full text] [doi: 10.1177/1088868310377394] [Medline: 20716644]
29. Pu L, Moyle W, Jones C, Todorovic M. The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. Gerontologist 2019 Jan 09;59(1):37-51. [doi: 10.1093/geront/gny046] [Medline: 29897445]
30. Choi M, Kong S, Jung D. Computer and internet interventions for loneliness and depression in older adults: a meta-analysis. Healthc Inform Res 2012 Sep;18(3):191-198 [FREE Full text] [doi: 10.4258/hir.2012.18.3.191] [Medline: 23115742]
31. Morris ME, Adair B, Miller K, Ozanne E, Hansen R, Pearce AJ, et al. Smart-Home Technologies to Assist Older People to Live Well at Home. Aging Sci 2013;01(01):1-9. [doi: 10.4172/2329-8847.1000101]
32. Morris ME, Adair B, Ozanne E, Kurowski W, Miller KJ, Pearce AJ, et al. Smart technologies to enhance social connectedness in older people who live at home. Australas J Ageing 2014 Sep;33(3):142-152. [doi: 10.1111/ajag.12154] [Medline: 24730370]
33. Pearce AJ, Adair B, Miller K, Ozanne E, Said C, Santamaria N, et al. Robotics to enable older adults to remain living at home. J Aging Res 2012 Dec 4;2012:538169 [FREE Full text] [doi: 10.1155/2012/538169] [Medline: 23304507]
34. Hagan R, Manktelow R, Taylor BJ, Mallett J. Reducing loneliness amongst older people: a systematic search and narrative review. Aging Ment Health 2014 Jan 17;18(6):683-693. [doi: 10.1080/13607863.2013.875122] [Medline: 24437736]
35. Miller KJ, Adair BS, Pearce AJ, Said CM, Ozanne E, Morris MM. Effectiveness and feasibility of virtual reality and gaming system use at home by older adults for enabling physical activity to improve health-related domains: a systematic review. Age Ageing 2014 Mar;43(2):188-195. [doi: 10.1093/ageing/aft194] [Medline: 24351549]
36. Medical Advisory Secretariat. Social isolation in community-dwelling seniors: an evidence-based analysis. Ont Health Technol Assess Ser 2008;8(5):1-49 [FREE Full text] [Medline: 23074510]
37. Shah SG, Nogueras D, van Woerden H, Kiparoglou V. Effectiveness of digital technology interventions to reduce loneliness in adults: a protocol for a systematic review and meta-analysis. BMJ Open 2019 Sep 27;9(9):e032455 [FREE Full text] [doi: 10.1136/bmjopen-2019-032455] [Medline: 31562164]
38. Higgins J, Thomas J, Chandler J, Cumpston M, Deeks J, Clarke M. Chapter I: Introduction. Cochrane Handbook for Systematic Reviews of Interventions Version 6.1 (Updated September 2020). 2020. URL: https://training.cochrane.org/handbook/archive/v6.1/chapter-i [accessed 2020-09-17]
39. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
40. Shah S, Nogueras D, van Woerden H, Kiparoglou V. The effectiveness of digital technology interventions to reduce loneliness in adult people: a systematic review and meta-analysis. PROSPERO 2019 CRD42019131524. 2019. URL: https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=131524 [accessed 2019-06-10]
41. What are some everyday examples of digital technology? Quora. URL: https://www.quora.com/What-are-digital-technologies [accessed 2019-06-06]

42. Hemberg J, Fischer RS. A window toward the world: older adults' experiences of becoming in health and developing as human beings through interacting with others using real video communication. Holist Nurs Pract 2018;32(2):90-97. [doi: 10.1097/HNP.0000000000000254] [Medline: 29432355]
43. Gao Q, Ebert D, Chen X, Ding Y. Design of a mobile social community platform for older Chinese people in urban areas. Hum Factors Man 2012 Jun 27;25(1):66-89. [doi: 10.1002/hfm.20523]
44. Austin J, Dodge HH, Riley T, Jacobs PG, Thielke S, Kaye J. A smart-home system to unobtrusively and continuously assess loneliness in older adults. IEEE J Transl Eng Health Med 2016 Jun 10;4:1-11. [doi: 10.1109/jtehm.2016.2579638]
45. Czaja SJ, Boot WR, Charness N, Rogers WA, Sharit J. Improving social support for older adults through technology: findings from the PRISM randomized controlled trial. Gerontologist 2018 May 8;58(3):467-477 [FREE Full text] [doi: 10.1093/geront/gnw249] [Medline: 28201730]
46. Tsai H, Tsai Y, Wang H, Chang Y, Chu HH. Videoconference program enhances social support, loneliness, and depressive status of elderly nursing home residents. Aging Ment Health 2010 Nov;14(8):947-954. [doi: 10.1080/13607863.2010.501057] [Medline: 21069600]
47. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Pers Soc Psychol 1980 Sep;39(3):472-480. [Medline: 7431205]
48. van der Heide LA, Willems CG, Spreeuwenberg MD, Rietman J, de Witte LP. Implementation of CareTV in care for the elderly: the effects on feelings of loneliness and safety and future challenges. Technol Disab 2012 Dec 12;24(4):283-291. [doi: 10.3233/tad-120359]
49. Gierveld JD, Tilburg TV. A 6-item scale for overall, emotional, and social loneliness. Res Aging 2006 Sep 11;28(5):582-598. [doi: 10.1177/0164027506289723]
50. Larsson E, Padyab M, Larsson-Lund M, Nilsson I. Effects of a social internet-based intervention programme for older adults: an explorative randomised crossover study. Br J Occup Ther 2016 Jul 19;79(10):629-636. [doi: 10.1177/0308022616641701]
51. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
52. Engelberg E, Sjöberg L. Emotional intelligence and inter-personal skills. In: Schulze R, Roberts RD, editors. Emotional Intelligence: An International Handbook. Cambridge, MA: Hogrefe & Huber; 2005:289-305.
53. Morton TA, Wilson N, Haslam C, Birney M, Kingston R, McCloskey L. Activating and guiding the engagement of seniors with online social networking. J Aging Health 2018 Jan;30(1):27-51. [doi: 10.1177/0898264316664440] [Medline: 27530332]
54. Jarvis MA, Padmanabhanunni A, Chipps J. An evaluation of a low-intensity cognitive behavioral therapy mHealth-supported intervention to reduce loneliness in older people. Int J Environ Res Public Health 2019 Apr 11;16(7):1305 [FREE Full text] [doi: 10.3390/ijerph16071305] [Medline: 30979042]
55. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Br Med J 2015 Jan 02;350:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
56. Munn Z, Tufanaru C, Aromataris E. JBI's systematic reviews: data extraction and synthesis. Am J Nurs 2014 Jul;114(7):49-54. [doi: 10.1097/01.NAJ.0000451683.66447.89] [Medline: 25742352]
57. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol 2018 Jan 10;18:5 [FREE Full text] [doi: 10.1186/s12874-017-0468-4] [Medline: 29316881]
58. Larsson E, Nilsson I, Lund ML. Participation in social internet-based activities: five seniors' intervention processes. Scand J Occup Ther 2013 Nov 30;20(6):471-480. [doi: 10.3109/11038128.2013.839001] [Medline: 24074300]
59. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide: Joanna Briggs Institute (JBI); 2020. [doi: 10.46658/JBIMES-20-04]
60. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions - Version 5.1.0. The Cochrane Collaboration. 2011. URL: https://handbook-5-1.cochrane.org/front_page.htm [accessed 2019-05-13]
61. Takeshima N, Sozu T, Tajika A, Ogawa Y, Hayasaka Y, Furukawa TA. Which is more generalizable, powerful and interpretable in meta-analyses, mean difference or standardized mean difference? BMC Med Res Methodol 2014 Feb 21;14:30 [FREE Full text] [doi: 10.1186/1471-2288-14-30] [Medline: 24559167]
62. Valtorta NK, Kanaan M, Gilbody S, Hanratty B. Loneliness, social isolation and social relationships: what are we measuring? A novel framework for classifying and comparing tools. BMJ Open 2016 Apr 18;6(4):e010799 [FREE Full text] [doi: 10.1136/bmjopen-2015-010799] [Medline: 27091822]
63. de Jong-Gierveld J, Kamphuls F. The development of a Rasch-Type Loneliness Scale. Appl Psychol Meas 2016 Jul 27;9(3):289-299. [doi: 10.1177/014662168500900307]
64. Eccleston C, Fisher E, Craig L, Duggan GB, Rosser BA, Keogh E. Psychological therapies (internet-delivered) for the management of chronic pain in adults. Cochrane Database Syst Rev 2014 Feb 26;2014(2):CD010152. [doi: 10.1002/14651858.CD010152.pub2] [Medline: 24574082]

65. Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Fam Pract 2000 Feb;17(Suppl 1):S11-S16. [doi: 10.1093/fampra/17.suppl_1.s11] [Medline: 10735262]
66. Review Manager (RevMan) Version 5. The Cochrane Collaboration. URL: https://community.cochrane.org/help/tools-and-software/revman-5 [accessed 2019-07-19]
67. Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to Meta-Analysis. Hoboken, New Jersey, United States: John Wiley & Sons; 2009:1-452.
68. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 2010 Apr 21;1(2):97-111. [doi: 10.1002/jrsm.12] [Medline: 26061376]
69. da Costa BR, Juni P. Systematic reviews and meta-analyses of randomized trials: principles and pitfalls. Eur Heart J 2014 Dec 14;35(47):3336-3345. [doi: 10.1093/eurheartj/ehu424] [Medline: 25416325]
70. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol 2011 Dec;64(12):1311-1316. [doi: 10.1016/j.jclinepi.2011.06.004] [Medline: 21802902]
71. Sedgwick P. Meta-analyses: how to read a funnel plot. Br Med J 2013 Mar 01;346(mar01 2):f1342. [doi: 10.1136/bmj.f1342]
72. Sutton A, Abrams K, Jones D, Sheldon T, Song F. Methods for Meta-Analysis in Medical Research. Hoboken, New Jersey, United States: John Wiley & Sons; 2000:1-348.
73. Noone C, McSharry J, Smalle M, Burns A, Dwan K, Devane D, et al. Video calls for reducing social isolation and loneliness in older people: a rapid review. Cochrane Database Syst Rev 2020 May 21;5(5):CD013632. [doi: 10.1002/14651858.CD013632] [Medline: 32441330]
74. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey, United States: Wiley-Blackwell; 2019:1-728.
75. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
76. Gierveld JD, Van Tilburg T. The De Jong Gierveld short scales for emotional and social loneliness: tested on data from 7 countries in the UN generations and gender surveys. Eur J Ageing 2010 Jun;7(2):121-130 [FREE Full text] [doi: 10.1007/s10433-010-0144-6] [Medline: 20730083]
77. van Staden WC, Coetzee K. Conceptual relations between loneliness and culture. Curr Opin Psychiatry 2010 Nov;23(6):524-529. [doi: 10.1097/YCO.0b013e32833f2ff9] [Medline: 20838346]
78. Barreto M, Victor C, Hammond C, Eccles A, Richins MT, Qualter P. Loneliness around the world: Age, gender, and cultural differences in loneliness. Pers Individ Dif 2021 Feb 01;169:110066 [FREE Full text] [doi: 10.1016/j.paid.2020.110066] [Medline: 33536694]
79. Heu LC, van Zomeren M, Hansen N. Does loneliness thrive in relational freedom or restriction? The culture-loneliness framework. Rev Gen Psychol 2020 Sep 24;25(1):60-72. [doi: 10.1177/1089268020959033]
80. Anderson G, Thayer C. A national survey of adults 45 and older: loneliness and social connections. AARP Foundation, Washington DC. 2018. URL: https://www.aarp.org/content/dam/aarp/research/surveys_statistics/life-leisure/2018/loneliness-social-connections-2018.doi.10.26419-2Fres.00246.001.pdf [accessed 2020-09-21]
81. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
82. Kushlev K, Proulx JD, Dunn EW. Digitally connected, socially disconnected: the effects of relying on technology rather than other people. Comput Hum Behav 2017 Nov;76:68-74. [doi: 10.1016/j.chb.2017.07.001]
83. Poscia A, Stojanovic J, La MD, Duplaga M, Grysztar M, Moscato U, et al. Interventions targeting loneliness and social isolation among the older people: an update systematic review. Exp Gerontol 2018 Feb;102:133-144. [doi: 10.1016/j.exger.2017.11.017] [Medline: 29199121]
84. Eccles AM, Qualter P. Review: Alleviating loneliness in young people - a meta-analysis of interventions. Child Adolesc Ment Health 2021 Feb;26(1):17-33. [doi: 10.1111/camh.12389] [Medline: 32406165]
85. Ballantyne A, Trenwith L, Zubrinich S, Corlis M. 'I feel less lonely': what older people say about participating in a social networking website. Qual Ageing Older Adults 2010 Sep 13;11(3):25-35. [doi: 10.5042/qiaoa.2010.0526]
86. Kahlbaugh PE, Sperandio AJ, Carlson AL, Hauselt J. Effects of playing wii on well-being in the elderly: physical activity, loneliness, and mood. Act Adapt Aging 2011 Oct;35(4):331-344. [doi: 10.1080/01924788.2011.625218]
87. Blažun H, Saranto K, Rissanen S. Impact of computer training courses on reduction of loneliness of older people in Finland and Slovenia. Comput Hum Behav 2012 Jul;28(4):1202-1212. [doi: 10.1016/j.chb.2012.02.004]
88. Blusi M, Asplund K, Jong M. Older family carers in rural areas: experiences from using caregiver support services based on Information and Communication Technology (ICT). Eur J Ageing 2013 Sep;10(3):191-199 [FREE Full text] [doi: 10.1007/s10433-013-0260-1] [Medline: 28804294]
89. Chopik WJ. The benefits of social technology use among older adults are mediated by reduced loneliness. Cyberpsychol Behav Soc Netw 2016 Sep;19(9):551-556 [FREE Full text] [doi: 10.1089/cyber.2016.0151] [Medline: 27541746]
90. Siniscarco MT, Love-Williams C, Burnett-Wolle S. Video conferencing: an intervention for emotional loneliness in long-term care. Act Adapt Aging 2017 Jun 15;41(4):316-329. [doi: 10.1080/01924788.2017.1326763]

91. Barbosa NB, Franz R, Judges R, Beermann C, Baecker R. Can digital technology enhance social connectedness among older adults? A feasibility study. J Appl Gerontol 2019 Jan;38(1):49-72. [doi: 10.1177/0733464817741369] [Medline: 29166818]
92. Hughes M. Loneliness and social support among lesbian, gay, bisexual, transgender and intersex people aged 50 and over. Ageing Soc 2015 Jul 16;36(9):1961-1981. [doi: 10.1017/s0144686x1500080x]

Abbreviations

DTI: digital technology intervention
GRADE: Grading of Recommendations Assessment, Development and Evaluation
PICO: participant, intervention, comparator, and outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SMD: standardized mean difference
UCLA: University of California, Los Angeles

Edited by R Kukafka; submitted 01.10.20; peer-reviewed by L Jordan, M Ji, J Hwang; comments to author 27.10.20; revised version received 09.11.20; accepted 19.04.21; published 04.06.21.

Please cite as:
Shah SGS, Nogueras D, van Woerden HC, Kiparoglou V
Evaluation of the Effectiveness of Digital Technology Interventions to Reduce Loneliness in Older Adults: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(6):e24712
URL: https://www.jmir.org/2021/6/e24712
doi:10.2196/24712
PMID:34085942

Review

Personal Health Information Management Among Older Adults: Scoping Review

Malgorzata Kolotylo-Kulkarni[1], PhD; Deborah E Seale[2], PhD; Cynthia M LeRouge[3], PhD

[1]Department of Information Management & Business Analytics, College of Business & Public Administration, Drake University, Des Moines, IA, United States

[2]Department of Public Health, College of Health Sciences, Des Moines University, Des Moines, IA, United States

[3]Department of Information Systems & Business Analytics, College of Business, Florida International University, Miami, FL, United States

Corresponding Author:
Malgorzata Kolotylo-Kulkarni, PhD
Department of Information Management & Business Analytics
College of Business & Public Administration
Drake University
2507 University Ave
Des Moines, IA, 50311
United States
Phone: 1 5152712007
Email: malgorzata.kolotylo-kulkarni@drake.edu

Abstract

Background: Older adults face growing health care needs and could potentially benefit from personal health information management (PHIM) and PHIM technology. To ensure effective PHIM and to provide supportive tools, it is crucial to investigate the needs, challenges, processes, and tools used by this subpopulation. The literature on PHIM by older adults, however, remains scattered and has not provided a clear picture of what we know about the elements that play a role in older adults' PHIM.

Objective: The goal of our review was to provide a comprehensive overview of extant knowledge on PHIM by older adults, establish the status quo of research on this topic, and identify research gaps.

Methods: We carried out a scoping review of the literature from 1998 to 2020, which followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) framework. First, we executed a broad and structured search. We then carried out a qualitative analysis of papers pertinent to the topic taking into consideration the five elements of the patient work system as follows: (1) personal-level factors, (2) PHIM tasks, (3) tools used, (4) physical settings of PHIM activities, and (5) socio-organizational aspects.

Results: The review included 22 studies. Consolidated empirical evidence was related to all elements of the patient work system. Multiple personal factors affected PHIM. Various types of personal health information were managed (clinical, patient-generated, and general) and tools were used (electronic, paper-based, and others). Older adults' PHIM was intertwined with their surroundings, and various individuals participated. The largest body of evidence concerned personal factors, while findings regarding the physical environment of PHIM were scarce. Most research has thus far examined older adults as a single group, and scant attention has been paid to age subgroups.

Conclusions: Opportunities for further PHIM studies remain across all elements of the patient work system in terms of empirical, design science, or review work.

(J Med Internet Res 2021;23(6):e25236) doi:10.2196/25236

KEYWORDS

personal health information management; health information management; scoping review; information management; consumer health informatics; medical informatics; patient participation

Introduction

Personal health information management (PHIM) is a process that involves creating, seeking, organizing, and sharing personal health information (PHI) of individuals to be engaged in their lives and their health care [1-5]. Patients who can access and manage their PHI may be more empowered to partner in their care. Effective PHIM can facilitate, for instance, patients' knowledge of their conditions [6] or adherence to treatment protocols [7]. However, PHIM is often challenging due to, for instance, many sources of information, and although there are tools designed to support PHIM, they differ in their level of accessibility, advancement, and cost.

Health consumers who could greatly benefit from effective PHIM to help support their health care and well-being are older adults. Older adults represent a growing subpopulation with approximately 52.5 million people aged 65 years or older in 2018 (35% increase from 2008) in the United States, and the number is projected to almost double by 2060 [8]. For this population, effective PHIM is of utmost importance as older adults often exhibit high health care needs [9] and costs [10] and may experience a decline in emotional well-being due to their health status [11].

Differences exist among older adults in terms of their experiences related to their health and their health care needs, which may drive different PHIM requirements and digital preferences among subgroups of older adults. These differences often correspond to various age subgroups within the older adult population.

For instance, the lives of older adults at midlife are often in flux [12]. They are frequently caretakers within their family dynamics (caring for spouses, children, parents, grandchildren, and/or siblings) and thereby may be managing a large volume of health information. As this group of people move into an older adult phase, they may be working longer or undergoing life transitions, such as retiring, which requires changes in health insurance coverage.

These transitioning older adults may differ from elderly people in their adoption of health technologies [13,14]. Many older adults have multiple health conditions, as comorbidities increase with age [15,16], and older adults with increasing health challenges exhibit high health care utilization [17]. These issues contribute to creating vast amounts of health-related information. Further, the elderly subpopulation is often on a fixed income and must closely manage health care costs. While elderly people may have more time to focus on managing their health information, their health conditions and potential cognitive decline may interfere with their ability to handle PHIM [18].

To ensure effective PHIM, design functional PHIM technology, and enable policymakers to devise practice interventions for older adults, we need to understand older adults' PHIM practices. The amount of effort and focus that a patient needs to assign to treatment has been coined "patient work" [19]. Such work not only entails the specific activities performed, but also includes and is shaped by the environmental and contextual elements that surround those activities.

Extant research indicates that PHIM is a complex and multidimensional phenomenon, as exemplified in the patient work framework [20]. This framework, while integrating prior models (the work system [21] and the SEIPS model [22]), consolidates the elements that are embodied in or impact patient work as follows: (1) person-related factors, (2) tasks carried out, (3) factors related to the tools used and information managed, (4) characteristics of the physical environment, and (5) socio-organizational aspects [20]. Indeed, to fully understand the landscape of PHIM practices by older adults, research needs to extend beyond one perspective or aspect of PHIM, a single technology, a health condition, or a single group of older adults. Insights go beyond the findings of an isolated study.

While limited PHIM literature reviews that attempt to consolidate extant knowledge on the topic do exist, with each one from a different perspective [4,23-27], few have focused on older adults (a previous study is an example of this research [28]). More so, those reviews did not provide a system view of the various factors that play a role in older adults' PHIM. Prior reviews examining PHIM by older adults focused on their patient portal use [28], but did not examine older adults' PHIM practices at a comprehensive general level. Other reviews studied the literature on medication management from the perspective of informal caregivers of older adults [29]. Literature related to older adults thus remains fragmented, and there is a need for an overview of extant empirical evidence on PHIM by older adults, particularly in light of the heterogeneity of PHIM.

The purpose of this review was thus to provide a synopsis of knowledge on PHIM by older adults, determine the status quo of this research, and identify gaps in it. This literature review systematizes and consolidates current empirical evidence on the needs and challenges older adults face, the current PHIM practices they carry out, the tools and information that they use for PHIM, the environment in which they manage their PHI, and the different stakeholders with which they interact. Furthermore, this study explores extant findings in the literature concerning PHIM differences among age subgroups of older adults. In light of the growing importance of electronic PHIM tools, we focused on PHIM literature published in the past two decades.

Methods

Overview

Literature reviews are well-recognized for their potential contributions. They have been shown to help establish the status quo of the literature, support theory testing, determine research gaps, and develop theory [30]. Recently, research pointed out the need for more literature review work in the information systems discipline, noted its significance in the field, and proposed suggestions on how rigorous and fruitful reviews may be executed [30].

Scoping reviews are particularly effective in answering broader research questions, carrying out a wider literature search, and providing an overview of research on a given topic [31-35]. They are also useful when examining complex and heterogenous phenomena [35]. To describe research on PHIM by older adults, we have thus carried out a scoping review of the literature on this topic. We were guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist [34].

Literature Search Strategy

Initially, we carried out several preliminary literature search processes, which enabled us to decide on a set of keywords and databases for the search. Our study's final literature search process consisted of the following three parts: (1) systematic search using online databases, (2) citation analysis of the full papers found during the search, and (3) citation analysis of four literature reviews published since 2009 related to PHIM. The database search further consisted of the following two stages: (1) the main search using generic PHIM keywords and (2) a detailed search using keywords representing main PHIM tasks identified through initial coding and review of the literature. We also carried out a citation analysis of articles identified during the database search and citation analysis of previous literature reviews by screening the papers cited by those studies (forward citation using Google Scholar). This approach helped us maximize the recall of the articles relevant to the study. We performed the searches between October and November 2019, with an update search conducted in December 2020 and January 2021. The update included Academic Search Complete (replacing Academic Search Premier) and did not include ABI/INFORM due to limited accessibility.

We followed search criteria (Multimedia Appendix 1) established by us for a broad and structured search process to ensure that articles relevant to our research objective and research questions were included. The search criteria were established to balance viability with breadth and comprehensiveness [36]. We focused on research published from 1998 to 2020 to cover the past two decades in order to balance recency (particularly in light of the increasing role of electronic PHIM tools) and comprehensiveness of empirical findings. We began our work by examining research published over two decades and continued to add literature as our work emerged. Due to the nature of the phenomenon (ie, PHIM encompasses multiple elements, such as actors, tools, and technologies), we decided to review only literature that examined PHIM among older adults without a focus on a specific technology (personal health records [PHRs] or wearable devices) or other actors (eg, caregivers).

Furthermore, to ensure the quality of the empirical evidence found and to establish the status of the development of this stream of research, we focused only on papers published in peer-reviewed journals. However, in order to ensure we did not miss any relevant recent findings, which could have been presented at conferences but have not been published in journal outlets, we also looked for conference papers in the 2019-2020 period.

Two researchers determined the articles to be included for the review to warrant their meeting of the inclusion criteria and their cohesiveness. Any ambiguities concerning inclusion were discussed and resolved.

Analysis of the Literature

To review the literature identified during the search, we carried out a qualitative analysis by adopting coding schema according to the patient work system [20] and using Dedoose. We proposed the patient work framework [20] as a lens from which to organize and connect findings of isolated tasks and tools (technology and others) used by older adults into a system of "patient work." Carrying out our analysis from the perspective of this framework enabled us to provide a comprehensive and consolidated view of the research on older adults' PHIM. The lead author did all the coding.

Upon completing the analysis, we summarized (1) the descriptive information about the eligible studies and (2) significant findings extracted from the papers relevant to our research questions.

For the review, we included five papers that also examined the perspectives of older adults' caregivers. However, we only incorporated findings from older adults' responses. Discussions of PHIM carried out by caregivers who were also older adults were omitted if the participant's age was not verifiable. Results not clearly attributed to older adults in the papers were also not included in the review.

We also included papers that examined PHIM by older adults even if they examined younger adults, but only if they also examined subgroups among older adults. We included only findings relevant to older adults and the subgroups among them. This search criterion was included owing to a small number of papers specifically studying older adults aged 50 years or above and carrying out a subgroup analysis.

Results

Literature Search Results

As a result of the search, 87 papers were eligible for in-depth examination, and we concluded the search with 22 papers eligible for qualitative analysis. The flowchart indicating the results of the literature search process is presented in Multimedia Appendix 2.

Study Characteristics

The majority (n=15) of papers were published since 2015. Reviewed research has taken different directions and examined the topic from various perspectives. The papers reviewed were slightly dominated by studies adopting a qualitative approach (13 papers), and eight papers undertook quantitative methods. The studies primarily included interviews, focus groups, survey questionnaires, and other methods such as review of existing patient portals or clinic appointment observations.

Most papers (n=17) examined these topics exclusively from the older adults' perspective, although five studies also included the point of view of older adults' informal caregivers.

Most papers (n=16) focused on older adults as a single group and did not distinguish across age subgroups. Details of the six papers focusing on age subgroups are provided in the section Older Adult Subgroup Study Findings.

Overall, concerning the study's purpose, the papers spanned from PHIM behavior studies and older adults' views of PHI and use thereof to PHIM technology use (such as patient portals) by older adults.

The summary of key information for the reviewed papers is provided in Multimedia Appendix 3.

PHIM by Older Adults

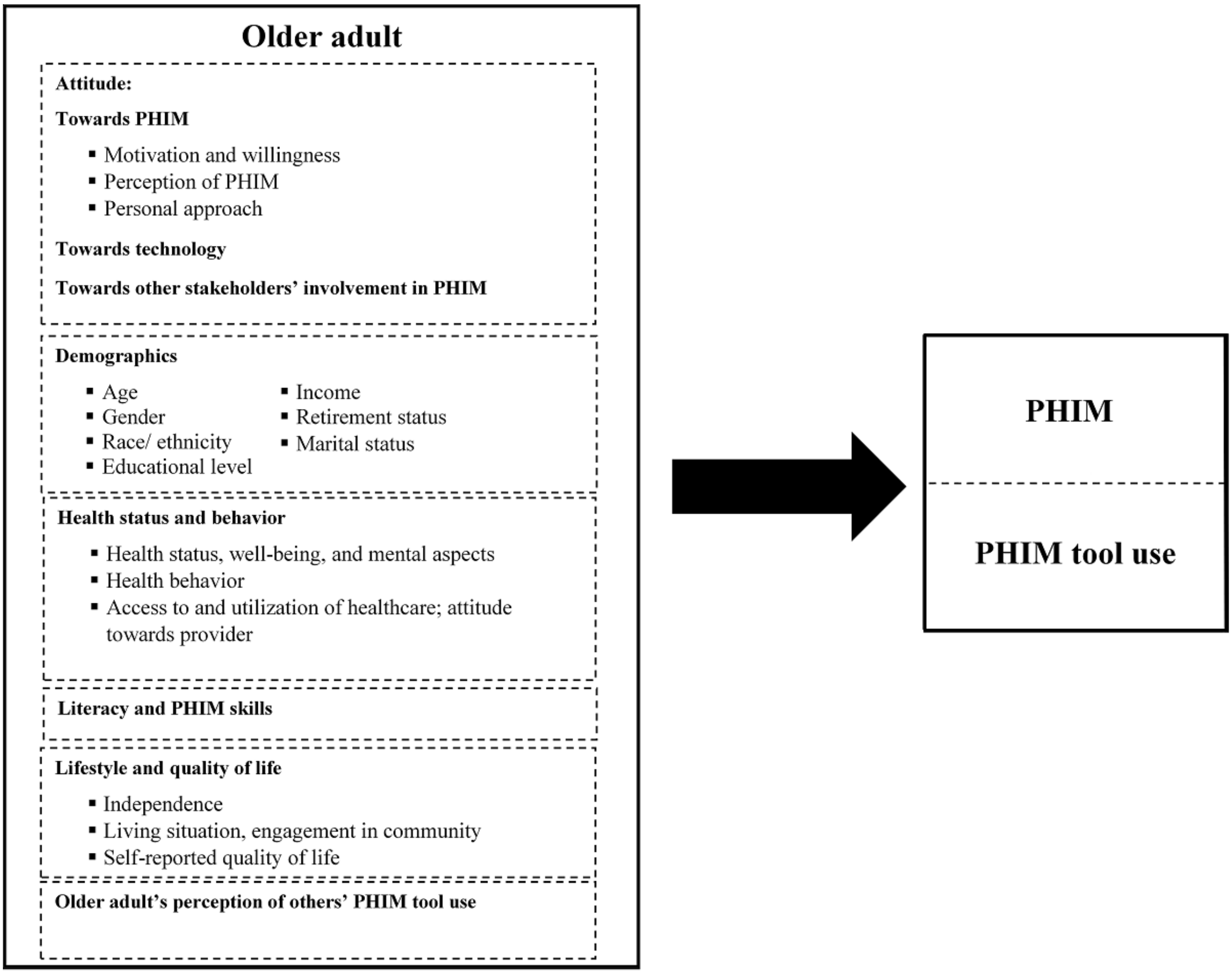
Below, we delineate the findings revealed in the literature concerning the elements of the patient work model [20] that play a role in older adults' PHIM.

Person-Related Factors That Drive or Challenge PHIM Among Older Adults

The reviewed literature showed that the major personal factors that drive or challenge older adults' PHIM span across their background and lifestyle. These factors include attitude toward PHIM [37], demographics [38], health status and behavior [39], literacy [40], lifestyle and quality of life [41], and perceptions of other stakeholders [42]. Many of these elements can vary in their effect on PHIM, as the literature has demonstrated differences and particular complexity when various aspects are studied (or from multiple perspectives). These disparities are exemplified, for instance, in the effects of gender, as some findings have shown that women are more likely to adopt online tools [43], while other findings have indicated that men exhibit more confidence in PHR use [44].

Figure 1 delineates these factors, while Multimedia Appendix 4 provides further details on them.

Figure 1. Person-related factors affecting personal health information management (PHIM) and PHIM tool use.



PHIM Tasks Carried Out by Older Adults and Their Characteristics

Managing personal health information involves multiple tasks and is performed over many (not linear) stages.

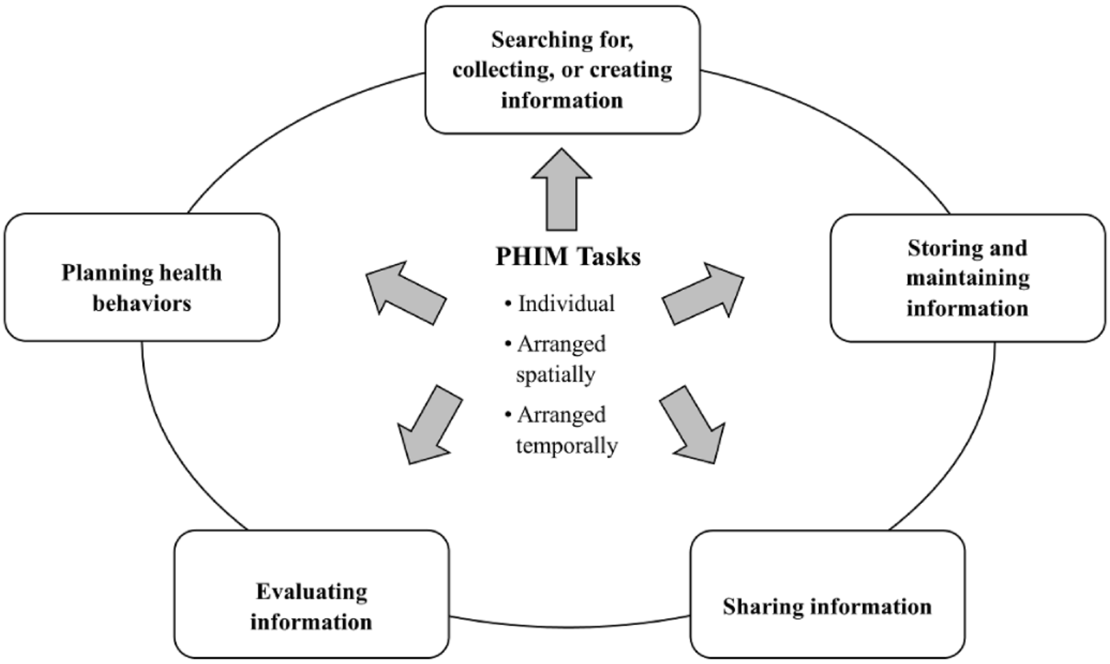
Older adults search for, collect, or create information [1,37,42,43,45-51]. They also share their PHI with others [1,39,42,43,45-49,51-53], make decisions concerning the storage and management of the information [1,43,45-47,51], and evaluate information [42,43,47,48], for instance, by reconciling conflicting information [47]. Importantly, PHIM tasks also include planning health behaviors with one's PHI.

Planning health behaviors include medication planning, such as filling pillboxes, purchasing medication, and planning how to keep medication; disposing of old medication; and ordering refills [41,48-50,52,54,55]. Another example is emergency planning, which has been noted as preparing or maintaining information for emergency situations [45,46,51]. These examples of planning as PHIM tasks particularly stand out owing to their predominance in the literature and the contextual nature of PHIM.

Figure 2 delineates the main PHIM tasks carried out by older adults. Further detailed findings on the tasks are provided in Multimedia Appendix 5 and Multimedia Appendix 6.

Figure 2 also shows the collective nature of PHIM tasks. First, PHIM tasks are highly individual, for instance, to what extent older adults are willing to share their PHI with others [39], and they vary across adults. For example, not everybody engages in various planning health behaviors, such as preparing emergency information [46].

Figure 2. Personal health information management (PHIM) tasks carried out by older adults.



PHIM tasks are also often synergistic with the environment in which they are executed. That is, tasks are intertwined and aligned with the location. For instance, older adults place pillboxes in various visible locations around the house to serve as reminders to take medication as a planned health behavior or choose to store their PHI record where it was originally generated (such as where blood pressure measurement is taken) [48]. Some older adults may choose to keep their nonprescription and prescription medication lists separate when reconciling differences between the two medication types [47].

Lastly, tasks are also temporally arranged, that is, tasks are entwined with one’s routine and other life activities. For instance, older adults may create information by checking their weight as part of their morning routine [48].

Personal Health Information Managed by Older Adults and the Types of Solutions or Tools That They Use to Support Their PHIM

Older adults manage various types of personal health information spanning clinical data [1,37,42,43,46,47,49], such as lab results [1,42,46,49]; patient-generated health data that includes clinical information [1,42,43,46-48,50,51,56], such as self-care logs [1,43,48,51], and information related to logistics and administration [1,42,46,51], such as emergency contact information [46,51]; and general health and wellness data, such as online information on medication side effects [42,47] and health educational materials [51]. Detailed findings on the information that older adults manage are shown in Multimedia Appendix 7.

Tools, solutions, and technologies that are currently offered or which older adults use to manage their PHI include electronic approaches [1,37,40,42,43,46-49,51,52], such as computers or laptops [42,46,51] and the internet [37,40,42-44,47-49,51,52]; paper-based approaches [1,37,41,42,46-48,51,54], such as printouts [37,42] and calendars [51,54]; and medical, every day, and other objects that include tangible objects [41,46-48,50,51,54,55], such as portable file cabinets [47] and pill boxes [41,48,50,51,54,55], and intangible objects [1,41,51], such as memory [1,41,51].

Detailed findings on the tools and methods older adults use for PHIM are provided in Multimedia Appendix 8 and Multimedia Appendix 9.

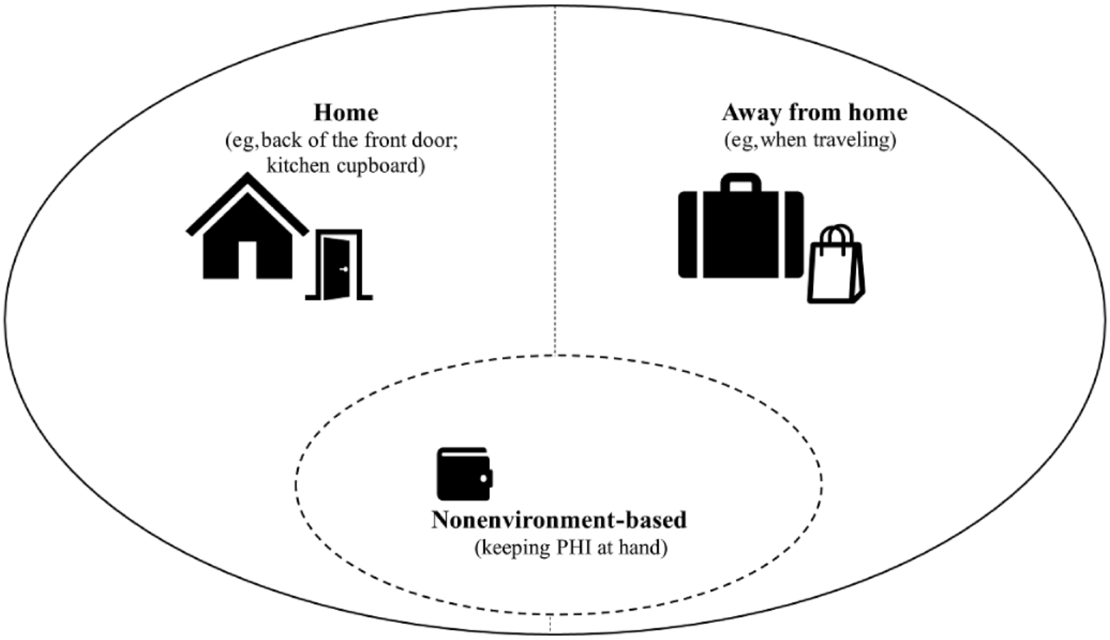
Physical Environments That Older Adults Occupy During PHIM and Their Characteristics

PHIM activities that are carried out by older adults occur in one’s house [1,46,55] and away from home [1,48]. Older adults use multiple locations in their homes for PHIM purposes, such as posting PHI on the back of their front door or fridge door

[1,46]. PHIM also crosses boundaries, as older adults, for instance, keep PHI at hand and carry it around (such as in their wallets) [46].

Figure 3 presents the physical environment of older adults' PHIM, and Multimedia Appendix 10 provides detailed findings in the literature on this aspect.

Figure 3. Physical environment of personal health information management by older adults. PHI: personal health information.

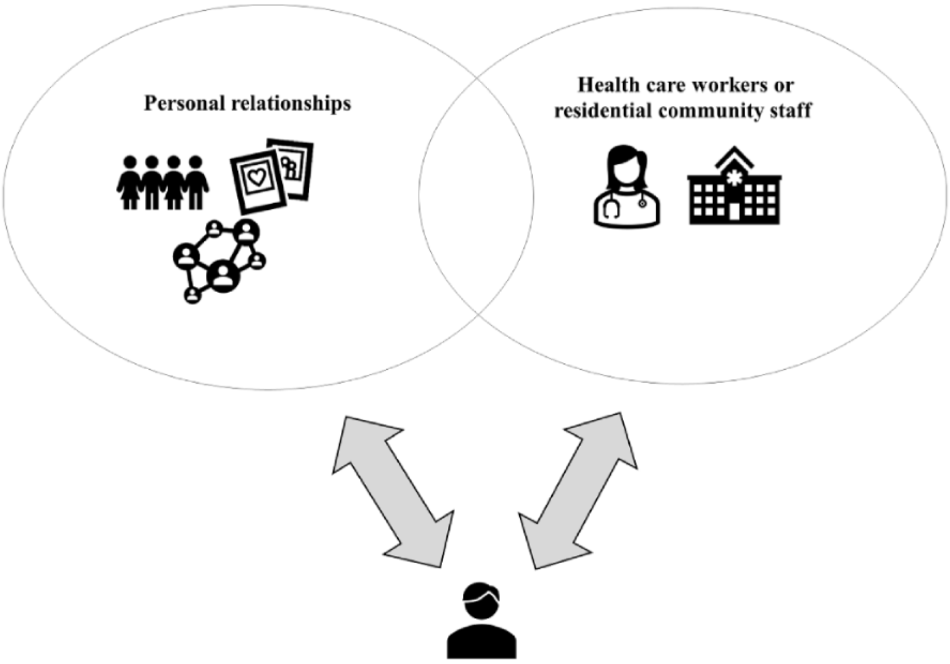


Socio-Organizational Environment in PHIM Among Older Adults: Stakeholders Involved

Many people are involved in older adults' PHIM and collaborate with them in different capacities to manage their PHI. These stakeholders include persons in the older adult's immediate circle (personal relationships), such as family, friends, and neighbors [1,37,39,41-44,46,47,51,53-56], and health care workers or retirement community staff, such as health care providers and professionals [1,37,42,43,46-48,50,56]. Sometimes, older adults particularly seek the help of their friends or relatives who have medical knowledge or expertise [42].

Figure 4 demonstrates the stakeholders with whom older adults interact during PHIM, and Multimedia Appendix 11 and Multimedia Appendix 12 delineate detailed literature findings on them.

Figure 4. Socio-organizational environment of older adult's personal health information management: stakeholders involved.



Older Adult Subgroup Study Findings

Table 1 presents a summary describing the papers that carried out an analysis based on age subgroups.

Six of the PHIM studies investigated older adult age subgroups (person factor) (Table 1). One examined a single older adult subgroup [41] and five compared two or more older adult subgroups [37,40,44,51,52]. As distinguishing among age subgroups was not their primary focus, two studies only reported one finding each related to age subgroup differences [40,51].

The number of older adult subgroups (person factor) studied and the age ranges of subgroups varied (5-year and 10-year increments, generational, median split, and very old) across the six studies (Table 1). Nevertheless, the findings were relatively consistent for the youngest and eldest of older adults regarding PHIM tasks, tools, and socio-organizational environmental factors. Physical environmental factors were notably absent from age subgroup findings.

A medication study [41] and one of the medical record studies [37] found that the eldest of older adults perceived the effective management of clinical PHI tasks to be necessary for (1) remaining in their homes [41], (2) communicating with their providers [37], and (3) taking better care of their health [37]. At the same time, all but one [40] of the four medical record studies found that the eldest of older adults were the least likely to use digital records and the least prepared to manage clinical PHI using digital technologies [37,44,52].

The eldest of older adults were also more likely to perceive the need for assistance from stakeholders and tools (digital and nondigital). The eldest subgroups shared their medical records to allow others to participate in their care [37] and relied on personal and health community caregivers to help them plan PHIM [51] and manage PHIM tools, that is, pill dispensers [41] and digital health records [52].

In contrast, the youngest of older adults were more likely to use and be prepared to use digital records [37,44,52], but less likely to use medical records to involve the family in their care [37] and more likely to use medical records to care for their children [37]. The findings are mixed for the two studies that examined middle older adult subgroups [37,52].

Table 1. Older adult personal health information management studies with age subgroup findings.

Authors, year published	PHIM[a] focus	Data collection		Sample size	Number of subgroups	Age subgroups
		Period	Source			
Arcury et al, 2017 [40]	ePortal[b]	2014-16	Interview	200	4	55-59, 60-64, 65-69, and ≥70 years
Gordon & Hornbrook, 2016 [52]	ePortal	2013-14	Admin; Survey	231,084; 3660[c]	3	65-69, 70-74, and 75-79 years
Huvila et al, 2018 [37]	PHIM & paper medical record.	2012	Survey	354	3	<52, 52-66, and ≥67 years
Logue & Effken, 2012 [44]	ePHR[d]	~2009[e]	Survey	38	2	65-77 and 78-93 years
Turner et al, 2021[f] [51]	PHIM	5-year period[g]	Interview; survey	88; 38[h]	4	60-69, 70-79, 80-89, and 90-99 years
Westerbotn et al, 2008 [41]	PHIM & medication management	2005	Interview	25	1	85-97 years

[a] PHIM: personal health information management.

[b] ePortal: electronic portal.

[c] Administrative data from the patient ePortal used to determine portal use (n=231,084) and identify a sample for the survey (n=3660).

[d] ePHR: electronic personal health record.

[e] Data collection period unspecified. It was inferred from a sentence in the manuscript.

[f] Turner et al, 2021 was published online in 2020.

[g] Exact timeframe unspecified.

[h] Subset of interview participants (n=88) willing to be contacted for the feedback survey (n=38).

Discussion

Implications of the Study

Research at large has recognized the peculiarities of midlife in terms of physical health, cognitive function, and social role [12]. Accordingly, scholars recognize the disparities between older adults at midlife and elderly people in terms of information behavior (eg, health information seeking [57]). However, the literature that we reviewed has largely not considered these differences. Only six papers carried out age subgroup analyses and only one paper included in the review examined the differences in the PHIM practices of midlife and elderly older adult subgroups.

Furthermore, the studies that recognized differences across age subgroups among older adults (eg, older adults and elderly people) adopted various cutoff ages among the subgroups. The ambiguity in the cutoff age used to distinguish older adults and elderly people in the reviewed studies suggests that there is no generally accepted cutoff age. Lack of a clear cutoff age for these two subgroups challenges a systematic approach to research on these two groups.

While an absolute cutoff age for older population subgroups creates some challenge, we need to look to the nature and purpose of the study for a path forward and to connect the literature. Underlying much of the PHIM literature that recognizes different subgroups for an older population is recognition that people typically have different generational idiosyncrasies as well as health needs at different stages of life. Belonging to a given generation can, conceivably, affect the socio-cultural characteristics of health consumers, thus potentially influencing their practices and approaches to PHIM technology. For example, research has acknowledged the uniqueness of age subgroups among older adults. Specifically, the literature has recognized that older adults at midlife are at a pivotal time in their life; hence, they have been referred to as pivotal agers [58]. In our review, we sought an objective means to consider subgroups among older adults; hence, we assumed the cutoff as the retirement age. However, various factors (generation, life experiences, etc) could be considered here. Extant literature has shown various approaches, with some research, for instance, driving the split by year of birth [37].

Our review corroborates the role that the various patient work elements of the PHIM system play for health consumers, as has been suggested in prior research [20]. The patient work model [20] has been shown to be valuable in observing the factors from different life and environment areas. It is particularly useful here in drawing more attention to socio-organizational aspects that affect older adults' PHIM. PHIM is affected by and intertwined with one's personal life as well as physical and socio-organizational environment. Thus, the factors involved in or influencing PHIM should be considered together to create a system, especially for those older adults who have comorbidities, and should be customized to an adult's unique health status.

Our review also shows that older adults adopt a variety of tools to support their PHIM, whereby not only electronic but also paper-based solutions are still commonly used.

Extant research has shown the role of the various elements of the patient work model [20]. However, the elements have received differing levels of attention. Our findings indicate that most empirical evidence in the literature thus far concerns person-related factors and the least evidence pertains to the physical environment of PHIM.

The complex and multidimensional nature of PHIM caused the nature of the search process to be quite challenging. Studies were found in multiple academic domains, and it was difficult to obtain a holistic perspective of which papers should be included and excluded. Our evaluations of whether studies should be included in the review were somewhat ambiguous and challenging, and necessitated establishing clear and detailed inclusion and exclusion criteria. Similar difficulties have been reported previously [28]. Moreover, the lack of existence of PHIM as a Medical Subject Headings (MeSH) term and the inconsistent use of keywords across papers complicated the discovery process.

Our review adds to the extant PHIM research. Our investigation extends prior work, which discussed the challenges of PHIM [23]. Our review also adds to previous literature reviews on PHIM tools [25,28], by examining the various types of PHIM tools used by older adults and the information they manage. We also extended the findings of earlier work [29] by corroborating the role that caregivers and other stakeholders play in older adults' PHIM. We extended the results of prior work [4], which also examined PHIM through the patient work model [20]. Our review included findings through 2020 and took a distinct perspective by focusing on older adults. Our findings are also consistent with research on medication management by older adults, which also used the patient work framework perspective [59].

Our review showcases several potential avenues for future empirical or design science research related to the various patient work elements that play a role in older adults' PHIM. Further research is needed to examine the idiosyncratic characteristics and challenges of older adults at midlife and elderly people. Additionally, it would be valuable to extend this research by investigating specific PHIM tools and tailoring their design toward different age subgroups among older adults. Furthermore, scant evidence regarding the characteristics of PHIM tasks and PHIM location suggests the need to inspect the nature (ie, attributes) of PHIM activities carried out by older adults and the physical environment of such activities.

PHIM research can also be extended by examining, for instance, the nature of involvement of the socio-organizational environment in older adults' PHIM practices. For example, this may be accomplished by focusing on the viewpoint of other stakeholders involved in older adults' PHIM, such as caregivers and providers.

Limitations

The limitations of our review's findings pertain primarily to the possibility of omitting relevant papers and the limited scope of the findings presented.

First, limiting our review to research published in peer-reviewed journals over the last 22 years and conference proceedings published in the last 2 years could have resulted in omission of relevant findings.

During the search, we did not include keywords such as those reflecting all the different types of PHIM technologies (eg, activity monitoring), as our focus was on older adults' characteristics and PHIM practices. It is conceivable, though, that literature on specific PHIM tools, which were omitted this way, could have also included empirical evidence on older adults' PHIM. It is thus possible that not every tool type was discovered in our review.

Lastly, the challenges of paper identification (caused by, for instance, the complexity of the topic and the occurrence of publications in many areas, as delineated above) could have resulted in erroneous omission of papers.

Conclusions

This paper contributes to research by consolidating and systematizing fragmented evidence from the literature on PHIM by older adults and establishing the status quo of research in this area. Our review shows that older adults' PHIM constitutes a system of patient work. Extant literature on this topic has so far focused largely on the personal characteristics of older adults, and the least attention has been paid to the physical environment of their PHIM. Most of the reviewed research did not differentiated between midlife and elderly people. Additionally, our review suggests that this area of research is still fairly recent.

Our review may be valuable for practitioners. Policymakers, for instance, may take into account the personal factors and older adults' socio-organizational environment affecting PHIM identified in our review to potentially pinpoint areas that necessitate or could be facilitated by practice interventions or organizational support. Furthermore, policymakers may also consider the use of paper-based and electronic tools by older adults in the context of information blocking and patient portal utilization.

The findings of our review may also encourage developers to consider the individual elements of the PHIM system in design and acknowledge the interdependencies among them. Such recognition could make the design of PHIM tools, such as patient portals, more holistic, resulting in tools that support PHIM as a system.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable suggestions with regard to the review format and practical implications of our study. We would also like to thank the National Association of Productivity & Organizing Professionals for inspiring this literature review through a discussion on the need for organized evidence and additional research around personal health information management.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Parameters of the literature search process.
[DOCX File , 16 KB - jmir_v23i6e25236_app1.docx]

Multimedia Appendix 2
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram of the literature search process.
[DOCX File , 38 KB - jmir_v23i6e25236_app2.docx]

Multimedia Appendix 3
Personal health information management study summary.
[DOCX File , 20 KB - jmir_v23i6e25236_app3.docx]

Multimedia Appendix 4
Person-related aspects that play a role in personal health information management by older adults.
[DOCX File , 26 KB - jmir_v23i6e25236_app4.docx]

Multimedia Appendix 5
Personal health information management tasks carried out by older adults.
[DOCX File , 18 KB - jmir_v23i6e25236_app5.docx]

Multimedia Appendix 6
Characteristics of personal health information management tasks carried out by older adults.
[DOCX File , 14 KB - jmir_v23i6e25236_app6.docx]

Multimedia Appendix 7
Personal health information managed by older adults.
[DOCX File , 15 KB - jmir_v23i6e25236_app7.docx]

Multimedia Appendix 8
Tools used by older adults for the purposes of personal health information management.
[DOCX File , 15 KB - jmir_v23i6e25236_app8.docx]

Multimedia Appendix 9
Differences in personal health information management tools used across age subgroups among older adults.

[DOCX File , 15 KB - jmir_v23i6e25236_app9.docx]

Multimedia Appendix 10

Physical environment of personal health information management by older adults.

[DOCX File , 14 KB - jmir_v23i6e25236_app10.docx]

Multimedia Appendix 11

Stakeholders involved or who play a role in older adults' personal health information management.

[DOCX File , 14 KB - jmir_v23i6e25236_app11.docx]

Multimedia Appendix 12

Differences in the socio-organizational environment in personal health information management across age subgroups among older adults.

[DOCX File , 13 KB - jmir_v23i6e25236_app12.docx]

References

1. Hartzler AL, Osterhage K, Demiris G, Phelan EA, Thielke SM, Turner AM. Understanding views on everyday use of personal health information: Insights from community dwelling older adults. Inform Health Soc Care 2018 Sep;43(3):320-333 [FREE Full text] [doi: 10.1080/17538157.2017.1297815] [Medline: 28399725]
2. Pratt W, Unruh K, Civan A, Skeels M. Personal health information management. Commun. ACM 2006 Jan;49(1):51-55 [FREE Full text] [doi: 10.1145/1107458.1107490]
3. Civan A, Skeels MM, Stolyar A, Pratt W. Personal health information management: consumers' perspectives. AMIA Annu Symp Proc 2006:156-160 [FREE Full text] [Medline: 17238322]
4. Agarwal R, Khuntia J. Personal Health Information Management and the Design of Consumer Health Information Technology: Background Report. Agency for Healthcare Research and Quality. 2009. URL: https://digital.ahrq.gov/sites/default/files/docs/citation/09-0075-EF.pdf [accessed 2021-06-01]
5. Moen A, Brennan P. Health@Home: The Work of Health Information Management in the Household (HIMH): Implications for Consumer Health Informatics (CHI) Innovations. J Am Med Inform Assoc 2005 Nov 01;12(6):648-656 [FREE Full text] [doi: 10.1197/jamia.m1758]
6. Woods S, Schwartz E, Tuepker A, Press N, Nazi K, Turvey C, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013 Mar 27;15(3):e65 [FREE Full text] [doi: 10.2196/jmir.2356] [Medline: 23535584]
7. Ross S, Lin CT. 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]
8. 2019 Profile of Older Americans. Administration for Community Living. 2020. URL: https://acl.gov/sites/default/files/Aging%20and%20Disability%20in%20America/2019ProfileOlderAmericans508.pdf [accessed 2020-08-20]
9. Ageing and health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2020-09-10]
10. How do health expenditures vary across the population? Health System Tracker. URL: https://www.healthsystemtracker.org/chart-collection/health-expenditures-vary-across-population/ [accessed 2020-09-10]
11. Steptoe A, Deaton A, Stone A. Subjective wellbeing, health, and ageing. The Lancet 2015 Feb;385(9968):640-648 [FREE Full text] [doi: 10.1016/s0140-6736(13)61489-0]
12. Lachman ME, Teshale S, Agrigoroaei S. Midlife as a Pivotal Period in the Life Course: Balancing Growth and Decline at the Crossroads of Youth and Old Age. Int J Behav Dev 2015 Jan 01;39(1):20-31 [FREE Full text] [doi: 10.1177/0165025414533223] [Medline: 25580043]
13. LeRouge C, Tao D, Ohs J, Lach H, Jupka K, Wray R. Challenges and Opportunities with Empowering Baby Boomers for Personal Health Information Management Using Consumer Health Information Technologies: an Ecological Perspective. AIMS Public Health 2014;1(3):160-181 [FREE Full text] [doi: 10.3934/publichealth.2014.3.160] [Medline: 29546084]
14. LeRouge C, Van Slyke C, Seale D, Wright K. Baby boomers' adoption of consumer health technologies: survey on readiness and barriers. J Med Internet Res 2014 Sep 08;16(9):e200 [FREE Full text] [doi: 10.2196/jmir.3049] [Medline: 25199475]
15. Davis JW, Chung R, Juarez DT. Prevalence of comorbid conditions with aging among patients with diabetes and cardiovascular disease. Hawaii Med J 2011 Oct;70(10):209-213 [FREE Full text] [Medline: 22162595]
16. Divo MJ, Martinez CH, Mannino DM. Ageing and the epidemiology of multimorbidity. Eur Respir J 2014 Oct;44(4):1055-1068 [FREE Full text] [doi: 10.1183/09031936.00059814] [Medline: 25142482]
17. Zayas CE, He Z, Yuan J, Maldonado-Molina M, Hogan W, Modave F, et al. Examining Healthcare Utilization Patterns of Elderly Middle-Aged Adults in the United States. Proc Int Fla AI Res Soc Conf 2016 May;2016:361-366 [FREE Full text] [Medline: 27430035]
18. Czaja S, Sharit J. Designing Training and Instructional Programs for Older Adults. Boca Raton, FL: CRC Press; 2012.

19. Strauss AL, Maines DR. Continual Permutations of Action. New York, NY: Routledge; 1993.
20. Valdez R, Holden R, Novak L, Veinot T. Transforming consumer health informatics through a patient work framework: connecting patients to context. J Am Med Inform Assoc 2015 Jan;22(1):2-10 [FREE Full text] [doi: 10.1136/amiajnl-2014-002826] [Medline: 25125685]
21. National Research Council, Division of Behavioral and Social Sciences and Education, Board on Human-Systems Integration, Committee on the Role of Human Factors in Home Health Care. Health Care Comes Home: The Human Factors. Washington, DC: National Academies Press; 2011.
22. Holden R, Carayon P, Gurses A, Hoonakker P, Hundt A, Ozok A, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
23. Lahtiranta J. Current challenges of personal health information management. J of Systems and Info Tech 2009 Aug 14;11(3):230-243 [FREE Full text] [doi: 10.1108/13287260910983614]
24. Davis S, Roudsari A, Raworth R, Courtney K, MacKay L. Shared decision-making using personal health record technology: a scoping review at the crossroads. J Am Med Inform Assoc 2017 Jul 01;24(4):857-866 [FREE Full text] [doi: 10.1093/jamia/ocw172] [Medline: 28158573]
25. Zhang Y, Fleischmann KR, Gao J, Xie B. A systematic review of the literature on consumers' use of patient portals: Preliminary results. Proc. Assoc. Info. Sci. Tech 2016 Feb 24;52(1):1-4 [FREE Full text] [doi: 10.1002/pra2.2015.1450520100142]
26. Saparova D. Motivating, influencing, and persuading patients through personal health records: a scoping review. Perspect Health Inf Manag 2012;9:1f [FREE Full text] [Medline: 22783154]
27. Smolij K, Dun K. Patient health information management: searching for the right model. Perspect Health Inf Manag 2006 Dec 12;3:10 [FREE Full text] [Medline: 18066368]
28. Sakaguchi-Tang D, Bosold A, Choi Y, Turner A. Patient Portal Use and Experience Among Older Adults: Systematic Review. JMIR Med Inform 2017 Oct 16;5(4):e38 [FREE Full text] [doi: 10.2196/medinform.8092] [Medline: 29038093]
29. Gillespie R, Mullan J, Harrison L. Managing medications: the role of informal caregivers of older adults and people living with dementia. A review of the literature. J Clin Nurs 2014 Dec;23(23-24):3296-3308 [FREE Full text] [doi: 10.1111/jocn.12519] [Medline: 24354583]
30. Schryen G, Benlian A, Rowe F, Gregor SD, Larsen KR. Literature Reviews in IS Research: What Can Be Learnt from the Past and Other Fields? CAIS 2017;41:759-774 [FREE Full text] [doi: 10.17705/1cais.04130]
31. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
32. 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]
33. Sucharew H, Macaluso M. Progress Notes: Methods for Research Evidence Synthesis: The Scoping Review Approach. J Hosp Med 2019 Jul 01;14(7):416-418. [doi: 10.12788/jhm.3248] [Medline: 31251164]
34. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
35. 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]
36. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5(1):69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
37. Huvila I, Enwald H, Eriksson-Backa K, Hirvonen N, Nguyen H, Scandurra I. Anticipating ageing: Older adults reading their medical records. Information Processing & Management 2018 May;54(3):394-407 [FREE Full text] [doi: 10.1016/j.ipm.2018.01.007]
38. Kim T, Choi M. Older adults' willingness to share their personal and health information when adopting healthcare technology and services. Int J Med Inform 2019 Jun;126:86-94 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.03.010] [Medline: 31029268]
39. Crotty B, Walker J, Dierks M, Lipsitz L, O'Brien J, Fischer S, et al. Information Sharing Preferences of Older Patients and Their Families. JAMA Intern Med 2015 Sep;175(9):1492-1497 [FREE Full text] [doi: 10.1001/jamainternmed.2015.2903] [Medline: 26147401]
40. Arcury T, Quandt S, Sandberg J, Miller JD, Latulipe C, Leng X, et al. Patient Portal Utilization Among Ethnically Diverse Low Income Older Adults: Observational Study. JMIR Med Inform 2017 Nov 14;5(4):e47 [FREE Full text] [doi: 10.2196/medinform.8026] [Medline: 29138129]
41. Westerbotn M, Fahlström E, Fastbom J, Agüero-Torres H, Hillerås P. How do older people experience their management of medicines? J Clin Nurs 2008 Mar;17(5A):106-115 [FREE Full text] [doi: 10.1111/j.1365-2702.2007.02151.x] [Medline: 18298761]

42. Kim S, Fadem S. Communication matters: Exploring older adults' current use of patient portals. Int J Med Inform 2018 Dec;120:126-136 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.10.004] [Medline: 30409337]
43. Zettel-Watson L, Tsukerman D. Adoption of online health management tools among healthy older adults: An exploratory study. Health Informatics J 2016 Jun;22(2):171-183 [FREE Full text] [doi: 10.1177/1460458214544047] [Medline: 25149210]
44. Logue M, Effken J. An exploratory study of the personal health records adoption model in the older adult with chronic illness. Inform Prim Care 2012;20(3):151-169 [FREE Full text] [doi: 10.14236/jhi.v20i3.21] [Medline: 23710840]
45. Turner AM, Osterhage K, Hartzler A, Taylor JO, Demiris G. Personal Health Information Management Practices of Older Adults: One Size Does Not Fit All. Stud Health Technol Inform 2019 Aug 21;264:1995-1996. [doi: 10.3233/SHTI190752] [Medline: 31438446]
46. Turner A, Osterhage K, Loughran J, Painter I, Demiris G, Hartzler A, et al. Emergency information management needs and practices of older adults: A descriptive study. Int J Med Inform 2018 Mar;111:149-158 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.12.001] [Medline: 29425626]
47. Haverhals L, Lee C, Siek K, Darr C, Linnebur S, Ruscin J, et al. Older adults with multi-morbidity: medication management processes and design implications for personal health applications. J Med Internet Res 2011 Jun 29;13(2):e44 [FREE Full text] [doi: 10.2196/jmir.1813] [Medline: 21715286]
48. Mickelson RS, Willis M, Holden RJ. Medication-related cognitive artifacts used by older adults with heart failure. Health Policy Technol 2015 Dec 01;4(4):387-398 [FREE Full text] [doi: 10.1016/j.hlpt.2015.08.009] [Medline: 26855882]
49. Portz J, Bayliss E, Bull S, Boxer R, Bekelman D, 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]
50. Swanlund S. Successful cardiovascular medication management processes as perceived by community-dwelling adults over age 74. Appl Nurs Res 2010 Feb;23(1):22-29 [FREE Full text] [doi: 10.1016/j.apnr.2008.03.005] [Medline: 20122507]
51. Turner A, Taylor J, Hartzler A, Osterhage K, Bosold A, Painter I, et al. Personal health information management among healthy older adults: Varying needs and approaches. J Am Med Inform Assoc 2021 Feb 15;28(2):322-333 [FREE Full text] [doi: 10.1093/jamia/ocaa121] [Medline: 32827030]
52. Gordon N, Hornbrook M. Differences in Access to and Preferences for Using Patient Portals and Other eHealth Technologies Based on Race, Ethnicity, and Age: A Database and Survey Study of Seniors in a Large Health Plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
53. Francis S, Smith F, Gray N, Denham M. Partnerships between older people and their carers in the management of medication. Int J Older People Nurs 2006 Dec;1(4):201-207 [FREE Full text] [doi: 10.1111/j.1748-3743.2006.00032.x] [Medline: 20925764]
54. Lakey S, Gray S, Borson S. Assessment of Older Adults' Knowledge of and Preferences for Medication Management Tools and Support Systems. Ann Pharmacother 2009 May 26;43(6):1011-1019 [FREE Full text] [doi: 10.1345/aph.1l704]
55. Roux P, Verloo H, Santiago-Delefosse M, Pereira F. The spatial dimensions of medication management by home-dwelling older adults after hospital discharge. Health Place 2019 Nov;60:102230 [FREE Full text] [doi: 10.1016/j.healthplace.2019.102230] [Medline: 31634701]
56. Tomlinson J, Silcock J, Smith H, Karban K, Fylan B. Post-discharge medicines management: the experiences, perceptions and roles of older people and their family carers. Health Expect 2020 Dec;23(6):1603-1613 [FREE Full text] [doi: 10.1111/hex.13145] [Medline: 33063445]
57. Czaja S, Sharit J, Hernandez M, Nair S, Loewenstein D. Variability among older adults in Internet health information-seeking performance. Gerontechnology 2010;9(1):46-55. [doi: 10.4017/gt.2010.09.01.004.00]
58. Jones W. "Information, knowledge and successful aging". UW Information School. URL: https://ischool.uw.edu/events/2019/10/research-symposium-william-jones [accessed 2020-12-29]
59. Mickelson R, Holden R. Medication management strategies used by older adults with heart failure: A systems-based analysis. Eur J Cardiovasc Nurs 2018 Jun;17(5):418-428 [FREE Full text] [doi: 10.1177/1474515117730704] [Medline: 28901787]

Abbreviations

PHI: personal health information
PHIM: personal health information management
PHR: personal health record

Edited by R Kukafka; submitted 23.10.20; peer-reviewed by K Fadahunsi, K Peterson, M Zahmatkeshan, J van Delden; comments to author 13.11.20; revised version received 08.01.21; accepted 13.05.21; published 07.06.21.

Please cite as:
Kolotylo-Kulkarni M, Seale DE, LeRouge CM
Personal Health Information Management Among Older Adults: Scoping Review
J Med Internet Res 2021;23(6):e25236
URL: https://www.jmir.org/2021/6/e25236
doi:10.2196/25236
PMID:34096872

Review

Effectiveness and Acceptance of Technology-Based Psychological Interventions for the Acute Treatment of Unipolar Depression: Systematic Review and Meta-analysis

Moritz Köhnen[1], MSc; Levente Kriston[1], Dipl Psych, PhD; Martin Härter[1], Dipl Psych, MD, PhD; Harald Baumeister[2], Dipl Psych, PhD; Sarah Liebherz[1], Dipl Psych, PhD

[1]Department of Medical Psychology, University Medical Center Hamburg–Eppendorf, Hamburg, Germany

[2]Department for Clinical Psychology and Psychotherapy, University of Ulm, Ulm, Germany

Corresponding Author:
Moritz Köhnen, MSc
Department of Medical Psychology
University Medical Center Hamburg–Eppendorf
Martinistr 52, Building West 26
Hamburg,
Germany
Phone: 49 40 7410 57705
Email: m.koehnen@uke.de

Abstract

Background: Evidence on technology-based psychological interventions (TBIs) for the acute treatment of depression is rapidly growing. Despite extensive research in this field, there is a lack of research determining effectiveness and acceptance of TBIs considering different application formats in people with a formally diagnosed depressive disorder.

Objective: The goal of the review was to investigate the effectiveness and acceptance of TBIs in people with diagnosed depression with particular focus on application formats (stand-alone interventions, blended treatments, collaborative and/or stepped care interventions).

Methods: Studies investigating adults with diagnosed unipolar depressive disorders receiving any kind of psychotherapeutic treatment delivered (at least partly) by a technical medium and conducted as randomized controlled trials (RCTs) were eligible for inclusion. We searched CENTRAL (Cochrane Central Register of Controlled Trials; August 2020), MEDLINE, PsycINFO, PSYNDEX, CINAHL (January 2018), clinical trial registers, and sources of grey literature (January 2019). Two independent authors decided about study inclusion and extracted data. We performed random effects meta-analyses to synthesize the data.

Results: Database searches resulted in 15,546 records of which 78 completed studies were included. TBIs delivered as stand-alone interventions showed positive effects on posttreatment depression severity when compared to treatment as usual (SMD –0.44, 95% CI –0.73 to –0.15, k=10; I^2=86%), attention placebo (SMD –0.51, 95% CI –0.73 to –0.30; k=12; I^2=66%), and waitlist controls (SMD –1.01, 95% CI –1.23 to –0.79; k=19; I^2=73%). Superior long-term effects on depression severity were shown when TBIs were compared to treatment as usual (SMD –0.24, 95% CI –0.41 to –0.07; k=6; I^2=48%) attention placebo (SMD –0.23, 95% CI –0.40 to –0.07; k=7; I^2=21%) and waitlist controls (SMD –0.74, 95% CI –1.31 to –0.18; k=3; I^2=79%). TBIs delivered as blended treatments (providing a TBI as an add-on to face-to-face treatment) yielded beneficial effects on posttreatment depression severity (SMD –0.27, 95% CI –0.48 to –0.05; k=8; I^2=53%) compared to face-to-face treatments only. Additionally, TBIs delivered within collaborative care trials were more effective in reducing posttreatment (SMD –0.20, 95% CI –0.36 to –0.04; k=2; I^2=0%) and long-term (SMD –0.23, 95% CI –0.39 to –0.07; k=2; I^2=0%) depression severity than usual care. Dropout rates did not differ between the intervention and control groups in any comparison (all $P \geq .09$).

Conclusions: We found that TBIs are effective not only when delivered as stand-alone interventions but also when they are delivered as blended treatments or in collaborative care trials for people with diagnosed depression. Our results may be useful to inform routine care, since we focused specifically on different application formats, formally diagnosed patients, and the long-term effectiveness of TBIs.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42016050413; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42016050413

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-028042

(J Med Internet Res 2021;23(6):e24584) doi:10.2196/24584

KEYWORDS

internet; digital health; digital mental health; telephone; psychotherapy; depressive disorder; systematic review; meta-analysis; technology-based psychological interventions

Introduction

Depression is a common [1] and debilitating mental disorder for affected individuals (eg, experiencing difficulties in everyday life) [2] and society (eg, burden of disease caused by depression) [3]. There are many effective treatment options, especially psychotherapeutic and pharmacological treatments, for people diagnosed with unipolar depression [1,4]. Despite the high prevalence, burden, and presence of many effective treatment options, depression is still undertreated [5].

Technology-based psychological interventions (TBIs) are seen as promising tools to supplement mental health care [6]. TBIs comprise a heterogeneous group of interventions [7] that can be delivered in different clinical phases of depression management (eg, acute treatment, relapse prevention); within these phases, they can be distinguished concerning their application format: stand-alone interventions, blended treatments, collaborative and/or stepped care interventions. In line with the German guideline for unipolar depression [1], we defined acute treatment as the treatment of an acute/present unipolar depressive episode aiming to reduce symptom burden so that response or remission of patients may be achieved. This clinical phase is differentiated from continuation and maintenance treatment and relapse prevention, which aim to further stabilize (responded or remitted patients of the acute treatment) and prevent relapse (or recurrence of new episodes) in the long term among people being at high risk. Additionally, TBIs vary in technical aspects (eg, delivery via videoconferencing tools), amount of human support, and theoretical background of the intervention [7]. Due to considerable diversity among TBIs and extensive research efforts capturing effectiveness and acceptance of TBIs for the acute treatment phase [8-10], there is need to address important neglected issues concerning TBIs.

First, TBIs in depression have already been widely researched resulting in high-quality evidence [11], and certain moderators influencing the success of treatment have been identified (eg, guided TBIs result in lower dropout rates than unguided TBIs) [8]. However, guideline recommendations are still limited to the general effectiveness of specific TBIs (eg, computerized cognitive behavioral therapy [cCBT] [1,4]). Additionally, there is no systematic review examining the effectiveness and acceptance of TBIs in the acute treatment phase regarding different application formats, even though the evidence base is available [11]. TBIs can be delivered as stand-alone interventions (TBIs replacing face-to-face [f2f] treatment), as blended treatments (combining TBIs and f2f treatment), or as part of stepped (eg, TBIs are used as a low-threshold initial treatment option for people with mild-to-moderate depressive disorder) and/or collaborative care models (TBIs may be provided alongside different treatment components, such as a TBI offered in addition to a care manager and general practitioners' care; see section *Application Formats of TBIs* for details). Blended treatments are usually conducted within a superiority (providing a full TBI alongside a full f2f treatment) or noninferiority (replacing some elements of f2f treatment by providing a TBI instead) trial design addressing different research questions (dose-response research focus vs cost-utility focus). A recent initiative considering both patients and clinicians emphasized top 10 research priorities in digital mental health [12]. One priority was to determine how treatment outcomes can be maximized by combining treatment options (eg, psychotherapy) with digital mental health interventions (ie, blended treatments). Considering application formats is of interest from the perspective of patients and clinicians, as it may help to determine effectiveness and acceptance of TBIs in a more differentiated manner, which may be relevant to inform clinical practice.

Second, the vast majority of research syntheses in this field included mixed populations based on symptom severity cutoff scores or the presence of diagnoses, providing valuable information on the effectiveness of interventions. To the best of our knowledge, there is only one systematic review evaluating internet- and mobile-based interventions in people with formally diagnosed depression; however, it is limited to waitlist control group comparisons [13]. In light of a comprehensive evidence base for TBIs in acute treatment [11] and the necessity of diagnoses to initiate treatment in mental health care, we focused only on studies requiring diagnosis of depression with the aim of determining the effectiveness and acceptance of TBIs. Additionally, high-quality evidence (RCTs) in clinical samples with diagnosed depression is the preferred source of evidence for the development and updating of clinical treatment guidelines such as the German [1] and United Kingdom [4] guidelines for depression.

Finally, to date there is no clarity regarding whether treatment effects achieved by TBIs are stable over time, since most reviews have focused on posttreatment intervention effects and have not considered long-term outcome data (for example, Karyotaki et al [14]).

By focusing specifically on different application formats, on people diagnosed with depression, and on long-term effectiveness of TBIs, we hope to provide a comprehensive evidence base that may be more useful to inform routine care than already existing evidence syntheses.

In summary, our main aim is to investigate posttreatment and long-term effectiveness and acceptance of TBIs delivered to people with diagnosed depression in the acute treatment phase, addressing the following research questions:

1. How effective and acceptable are TBIs delivered as stand-alone interventions compared to f2f treatment, attention placebo, treatment as usual (TAU), waitlist and no-treatment controls, and other TBIs?
2. How effective and acceptable are TBIs delivered as blended treatments (TBI plus f2f treatment) compared to f2f treatment (including psychotherapy, medication, TAU)?
3. How effective and acceptable are TBIs delivered as stepped and/or collaborative care approaches compared to TAU?

Methods

The study was part of a larger research synthesis project (comparative effectiveness of Technology-Based Interventions in Different Steps of Depression Care [TIDECA]) that was prospectively registered with International Prospective Register of Systematic Reviews (PROSPERO) [CRD42016050413] and described in the study protocol published elsewhere [15].

Search Strategy

The search was not limited by date, language, or publication status. We contacted first authors of all included publications for additional information on further (un)published trials and specific study information (see Köhnen et al [15] for details on the literature search/strategy).

Selection Criteria

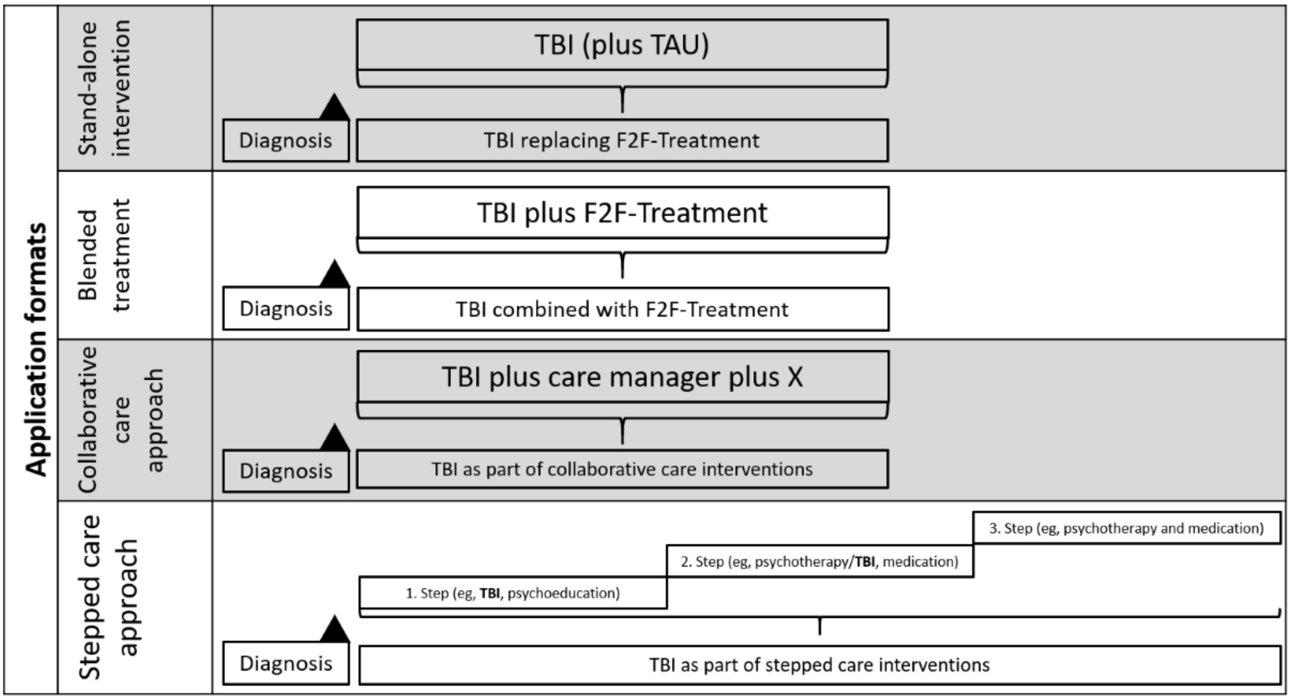
See study protocol [15] for more details on eligibility criteria. Our inclusion criteria were (1) at least 80% of sample having a diagnosed unipolar depression (assessed by criteria of a formal classification system or by conducting a diagnostic interview [eg, F32.x, F33.x, or F34.1 according to the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision*]) with any comorbidities in the acute treatment phase for depression and consisted of adults aged 18 years and older, (2) intervention was at least partly delivered through technical devices (eg, telephone, smartphone, computer), (3) intervention was based on an explicit psychotherapeutic theory, and (4) study was an individual or cluster RCT.

Our exclusion criteria were (1) participants were solely diagnosed by applying cutoff scores on symptom severity scales or when they had a depressive episode in the course of a bipolar disorder, (2) concurrent conditions (either somatic or mental) were the focus of the intervention, or (3) intervention provided solely psychoeducational content, patient decision aids, or depression management tools or focused exclusively on medication adherence.

Application Formats of TBIs

Since we placed a special focus on application formats in this review, they are presented visually in Figure 1. We applied a rather broad definition for blended treatments, since we included all studies that provided any type of f2f treatment tailored to depression (eg, psychotherapy, medication, depression specific general practitioner care) in addition to TBIs irrespective of the study's definition/label. In contrast, trials concurrently providing TAU in addition to TBIs were not considered blended treatments (but considered for the comparison TBI vs TAU) if TAU consisted of systematically offered generic treatments (eg, general practitioner care for all participants) that were not specifically tailored to depression. Since RCTs for blended treatment may be delivered in different designs (eg, superiority, noninferiority) resulting in content-related heterogeneity of interventions (eg, fewer therapeutic contacts), we decided to conduct meta-analyses separately.

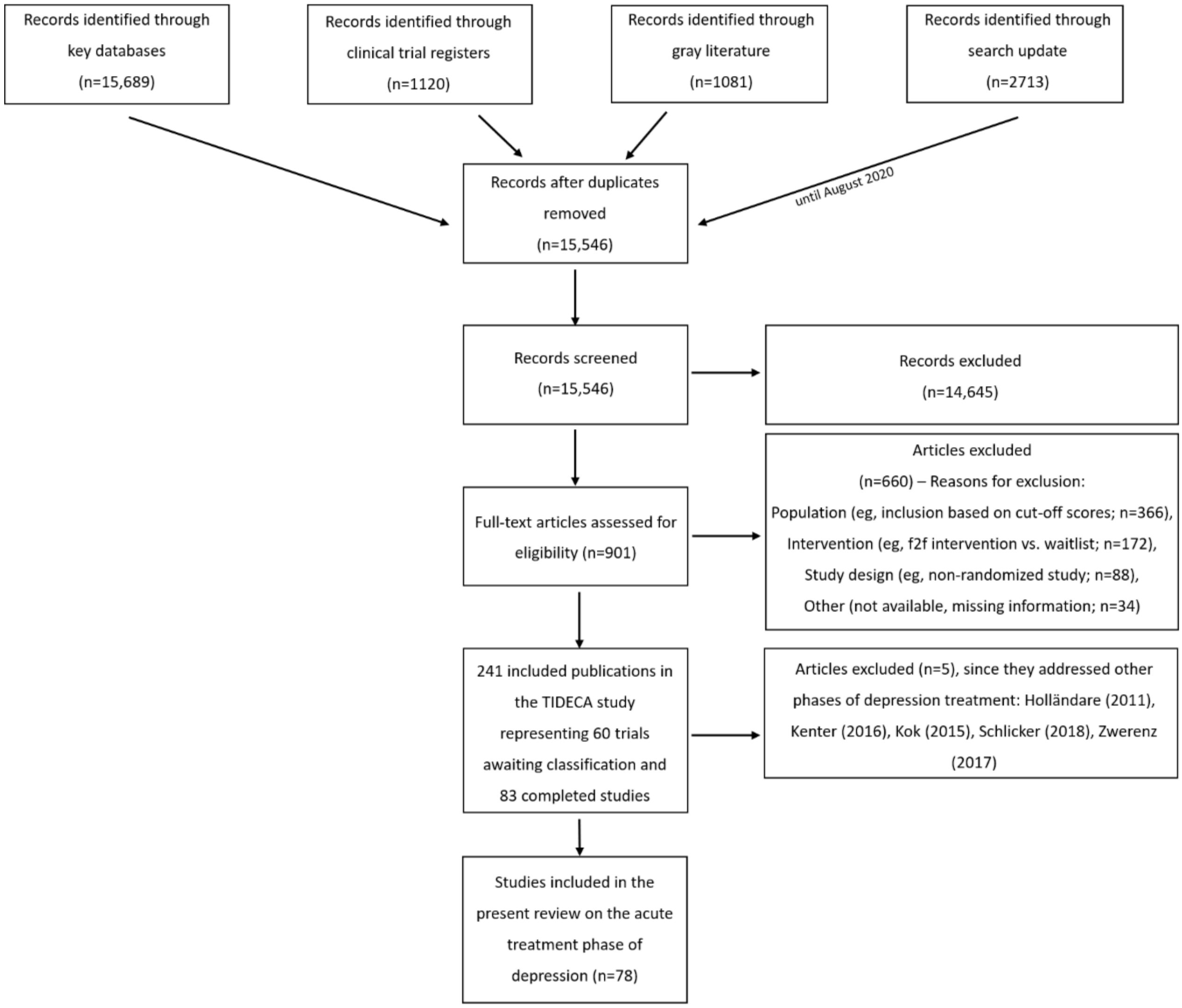
Figure 1. Illustration of potential application formats of technology-based psychological interventions.



Selection Procedure

The study flowchart is presented in Figure 2. Electronic searches yielded 20,603 records. After deduplication, 15,546 records were screened by title and abstract. Two reviewers (MK, SL) independently screened the first 100 records for inclusion. Since the interrater reliability for this sample was found to be high (98%), only one reviewer (MK) screened the remaining records in the course of the title/abstract screening. The second reviewer (SL) assessed publications labeled unclear by the first reviewer. Selected full-text articles (n=901) were subsequently assessed for inclusion by 2 independent reviewers (MK, MD). Discrepancies were resolved by discussion with a third reviewer (SL). In total, 241 publications representing 143 trials (83 completed studies and 60 ongoing studies awaiting further classification) fulfilled all inclusion criteria for the TIDECA study [11]. Of those, 78 completed studies assessed the acute treatment phase.

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart.



Data Extraction

See Köhnen et al [15] for detailed information on extracted data and extraction procedure.

Quality Appraisal

Risk of bias was independently assessed by 2 reviewers (from a group of 5 reviewers: MK, EW, MD, SL, TS) following Cochrane guidance (including the following domains for RCTs: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other bias) [16]. In line with a previous operationalization [17], we specified the domain *other bias* using the following 3 categories: insufficient treatment adherence, allegiance bias, and attention bias. Selective outcome reporting was categorized as *unclear risk* (trial registration or study protocol were missing or there was a deviation in one secondary outcome) or *high risk* (there were deviations in one primary or ≥2 secondary outcomes that could not be justified by the study authors). Disagreements were resolved by discussion or by consulting another reviewer (SL). Interrater reliability for risk of bias ratings was calculated to be 74%.

Data Analysis

Meta-analyses were computed applying random effects models [18] since we assumed that heterogeneity regarding the sample, treatment, and methodological features of the included studies would be best captured by assuming that moderately diverging study-specific effect estimates are distributed around a grand mean [19]. Results were visually displayed as forest plots.

Continuous data (posttreatment and long-term depression severity) were analyzed as standardized mean differences (SMDs). Dichotomous data ([any] dropouts) were analyzed using the risk ratio (RR). We calculated 95% confidence intervals for all estimates. In addition, we computed 95% prediction intervals (PIs) for meta-analysis (when possible) capturing the range in which the effect of a new study (in a different setting) is expected; PIs can be very imprecise when only a few studies are considered [20].

Studies with multiple treatment groups were considered by combining data from interventional study arms (ie, pooling of means and standard deviations for continuous data and summing up sample sizes and people with events for binary data) when possible to avoid a unit-of-analysis error [16].

In cases of missing or unclear data, we contacted the corresponding authors. Intention-to-treat (ITT) analyses were used when reported by the included studies. When ITT data were not reported, we used the analysis defined as primary by the authors of the trial. Data on dichotomous outcomes were excluded from data analysis if there were no events in either study arm, since the direction and magnitude of a potential effect is not indicated [16].

We assessed statistical heterogeneity in the included studies by using a Cochran Q test and quantified it using the I^2 statistic [21]. As defined in the study protocol [15], we considered I^2 values of 50% or more as indicators of relevant statistical heterogeneity requiring further exploration. If indicated, we explored heterogeneity either quantitatively by means of a priori (see Köhnen et al [15]) and post hoc subgroup analyses (if the number of studies was sufficient [≥10]) or narratively (if only a few studies were available [<10]). We tested for possible reporting biases and small-study effects using visual examination of funnel plots (when useful). Possible control interventions and comparisons of interests were prespecified in our protocol [15] and used to structure our results section. All meta-analyses were computed by using Review Manager 5.4 (Cochrane Collaboration); descriptive data (eg, mean age of included participants) and PIs were calculated using Excel 2013 (Microsoft Corp).

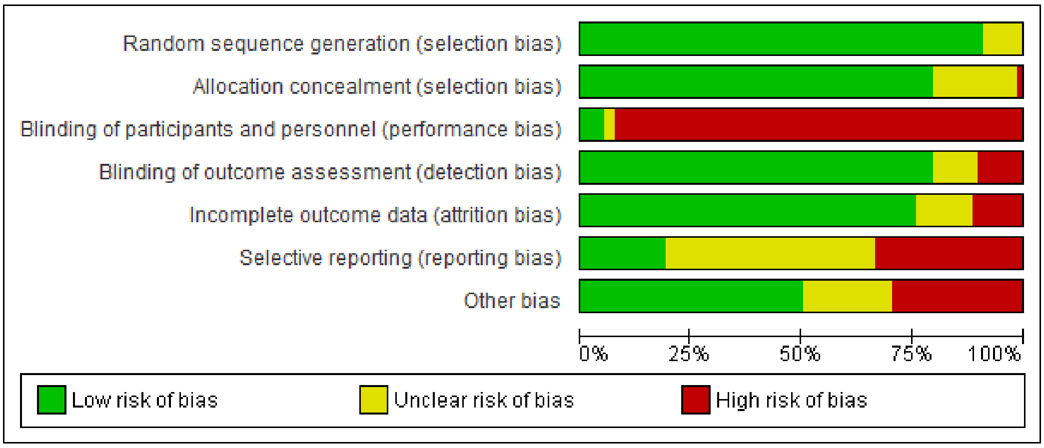
Results

A table summarizing all meta-analytic results can be found in Multimedia Appendix 1.

Study Characteristics and Quality of Included Studies

Overall, the selected studies (n=78) included 13,180 participants ranging from 14 to 1089 per study. The mean age of participants was 45.15 (SD 12.01) years, and two-thirds (8029/11981, 67.01%) were female. TBIs in the included studies were delivered as stand-alone interventions (61/78; 78%), blended treatments (12/78; 15%), collaborative care (3/78; 4%), or stepped care trials (2/78; 3%). Duration of TBIs ranged from 1 week to 52 weeks, with most interventions lasting between 6 weeks and 12 weeks (median treatment length of 8 weeks). Interventions of 8 weeks' duration were the most frequent (26/89; 29%) in the included studies (see Multimedia Appendix 2 [22-99] for baseline diagnoses). TBIs were based on 13 therapeutic rationales with most (83/101, 82.2%) based on CBT approaches (see Multimedia Appendix 3 for details). Concerning the applied technical medium, most TBIs were delivered via the internet (55/101, 54.5%), followed by telephone (12/101, 11.9%), offline computer programs (8/101, 7.9%), and videoconferencing tools (3/101, 3.0%). Additionally, 22.8% (23/101) of interventions applied more than one technical medium (internet-based treatment plus telephone support was most frequently [17/101, 16.8%] combined). The most common source of risk of bias was nonblinding of participants and personnel, selective reporting, and other bias (especially due to insufficient treatment adherence; Figure 3; see Multimedia Appendix 4 [22-99] for details).

Figure 3. Risk of bias assessment across included studies (n=78).



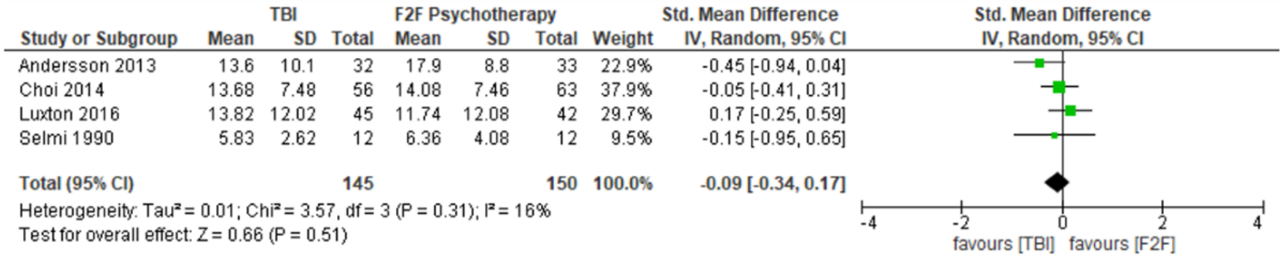
Stand-Alone Interventions

TBI Versus Face-to-Face Treatment

There were 6 RCTs comparing TBIs with f2f treatments [23,32,36,58,66,84]; 4 delivered therapist-administered treatment via videoconferencing [32,36,58] or telephone [66], and 2 delivered guided internet-based [23] or computer-based treatment [84]. There was no significant difference in posttreatment (SMD –0.09, 95% CI –0.34 to 0.17; I^2=16%; 95% PI –0.80 to 0.62) or long-term depression severity (2 months to 12 months; SMD –0.23, 95% CI –0.47 to 0.01; I^2=0%; 95% PI –0.76 to 0.3) between TBI and f2f interventions. There was no statistically significant difference in dropout rates between interventions (RR 0.85, 95% CI 0.63 to 1.15; I^2=17%; 95% PI 0.44 to 1.65; see Figure 4).

Figure 4. Forest plots on technology-based psychological intervention versus face-to-face-treatment.

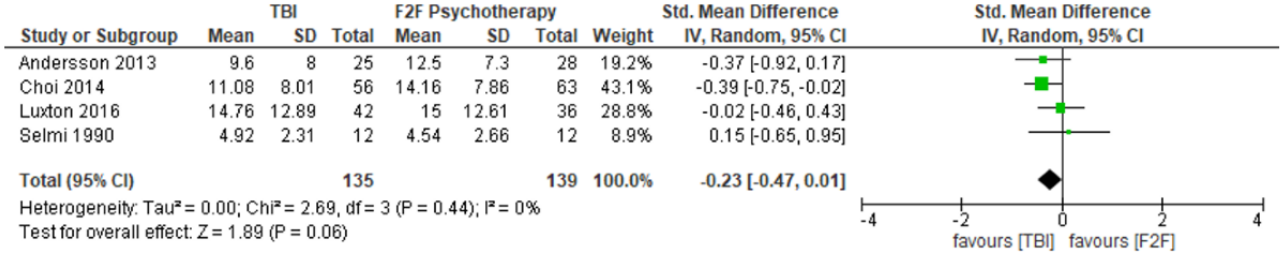
a) Post-treatment depression severity



	TBI			F2F Psychotherapy				Std. Mean Difference
Study or Subgroup	**Mean**	**SD**	**Total**	**Mean**	**SD**	**Total**	**Weight**	**IV, Random, 95% CI**
Andersson 2013	13.6	10.1	32	17.9	8.8	33	22.9%	-0.45 [-0.94, 0.04]
Choi 2014	13.68	7.48	56	14.08	7.46	63	37.9%	-0.05 [-0.41, 0.31]
Luxton 2016	13.82	12.02	45	11.74	12.08	42	29.7%	0.17 [-0.25, 0.59]
Selmi 1990	5.83	2.62	12	6.36	4.08	12	9.5%	-0.15 [-0.95, 0.65]
Total (95% CI)			**145**			**150**	**100.0%**	**-0.09 [-0.34, 0.17]**

Heterogeneity: Tau^2 = 0.01; Chi^2 = 3.57, df = 3 (P = 0.31); I^2 = 16%
Test for overall effect: Z = 0.66 (P = 0.51)

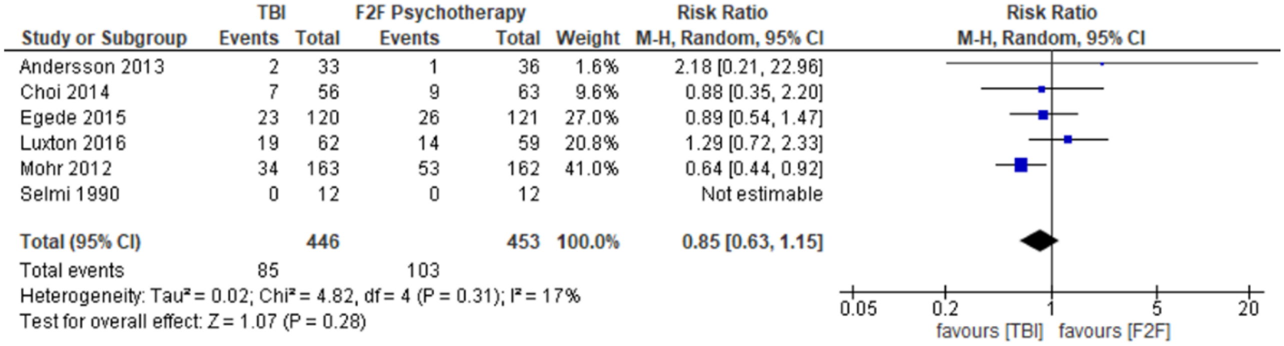
b) Long-term depression severity



	TBI			F2F Psychotherapy				Std. Mean Difference
Study or Subgroup	**Mean**	**SD**	**Total**	**Mean**	**SD**	**Total**	**Weight**	**IV, Random, 95% CI**
Andersson 2013	9.6	8	25	12.5	7.3	28	19.2%	-0.37 [-0.92, 0.17]
Choi 2014	11.08	8.01	56	14.16	7.86	63	43.1%	-0.39 [-0.75, -0.02]
Luxton 2016	14.76	12.89	42	15	12.61	36	28.8%	-0.02 [-0.46, 0.43]
Selmi 1990	4.92	2.31	12	4.54	2.66	12	8.9%	0.15 [-0.65, 0.95]
Total (95% CI)			**135**			**139**	**100.0%**	**-0.23 [-0.47, 0.01]**

Heterogeneity: Tau^2 = 0.00; Chi^2 = 2.69, df = 3 (P = 0.44); I^2 = 0%
Test for overall effect: Z = 1.89 (P = 0.06)

c) Dropout rates from treatment



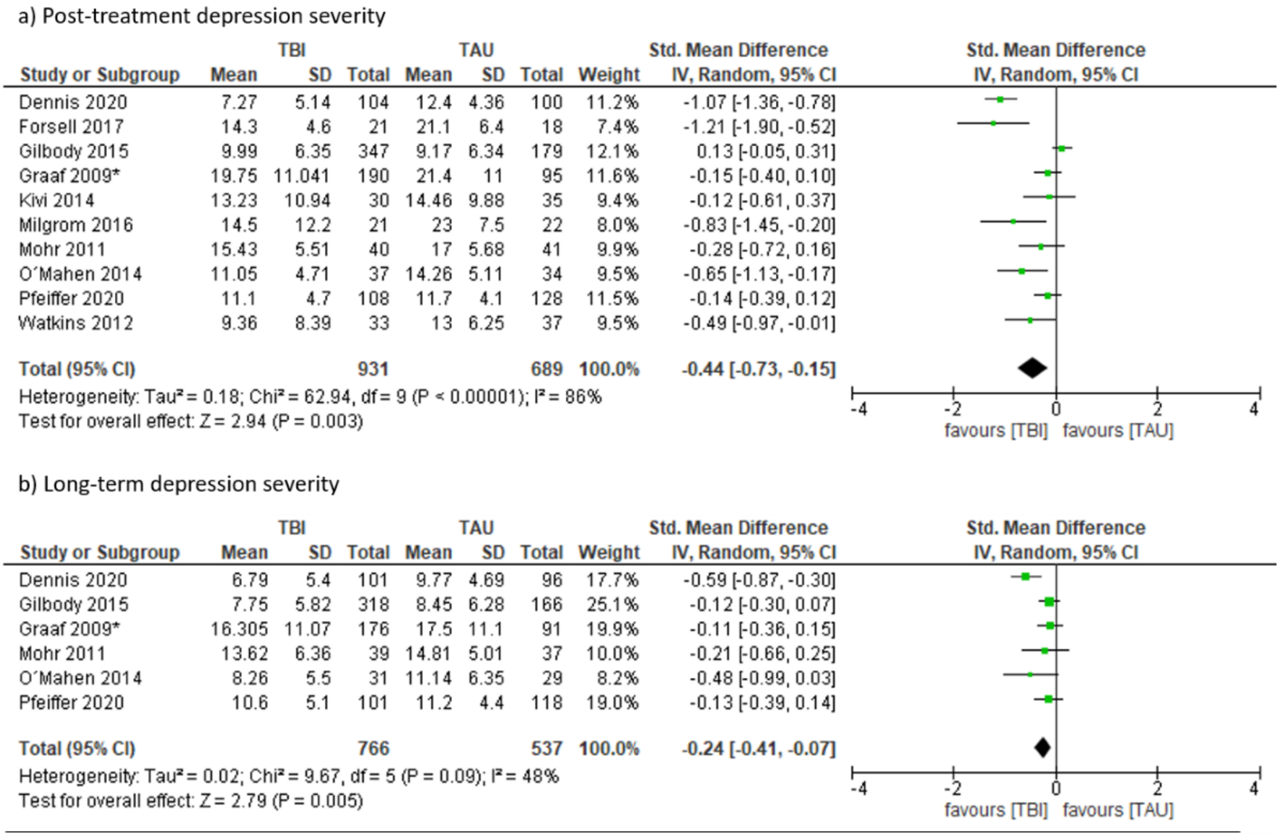
	TBI		F2F Psychotherapy			Risk Ratio
Study or Subgroup	**Events**	**Total**	**Events**	**Total**	**Weight**	**M-H, Random, 95% CI**
Andersson 2013	2	33	1	36	1.6%	2.18 [0.21, 22.96]
Choi 2014	7	56	9	63	9.6%	0.88 [0.35, 2.20]
Egede 2015	23	120	26	121	27.0%	0.89 [0.54, 1.47]
Luxton 2016	19	62	14	59	20.8%	1.29 [0.72, 2.33]
Mohr 2012	34	163	53	162	41.0%	0.64 [0.44, 0.92]
Selmi 1990	0	12	0	12		Not estimable
Total (95% CI)		**446**		**453**	**100.0%**	**0.85 [0.63, 1.15]**
Total events	85		103			

Heterogeneity: Tau^2 = 0.02; Chi^2 = 4.82, df = 4 (P = 0.31); I^2 = 17%
Test for overall effect: Z = 1.07 (P = 0.28)

TBI Versus Treatment as Usual

There were 12 RCTs testing TBIs against TAU [34,35,39-41,51,57,63,64,72,74,92], 8 of which explicitly stated that TAU was also administered in the TBI condition [34,35,39-41,57,74,92]. TBIs were delivered either with [39-41,51,63,72,74,92] or without [34,57] guidance or they were therapist-administered [35,64]. TAU consisted of care by a general practitioner [34,40,41,57,92], a heterogeneous mix of treatment options depending on resources and routines [51,63,72,74], care by community-based outpatient clinics and any non-Veterans Affairs facilities [64], and antenatal [39] or postpartum care [35]. Depression severity at posttreatment, with considerable heterogeneity (SMD –0.44, 95% CI –0.73 to –0.15; I^2=86%; 95% PI –1.48 to 0.60, and in the long term (6 months to 12 months; SMD –0.24, 95% CI –0.41 to –0.07; I^2=48%; 95% PI –0.70 to 0.22) was statistically significantly lower in the TBI condition (see Figure 5). Data on dropout rates were either not usable or missing. Prespecified subgroup analyses exploring heterogeneity for posttreatment depression severity were not conducted, as too few studies were available. Further exploration of heterogeneity did not reveal any specific source of variation. However, heterogeneity may be explained by the rather broad TAU condition, which consisted of various treatment options depending on the specific health care context where the intervention was delivered. Visual inspection of the funnel plot was not suspicious (Multimedia Appendix 5).

Figure 5. Forest plots on technology-based psychological intervention versus treatment as usual.



a) Post-treatment depression severity

Study or Subgroup	TBI Mean	TBI SD	TBI Total	TAU Mean	TAU SD	TAU Total	Weight	Std. Mean Difference IV, Random, 95% CI
Dennis 2020	7.27	5.14	104	12.4	4.36	100	11.2%	-1.07 [-1.36, -0.78]
Forsell 2017	14.3	4.6	21	21.1	6.4	18	7.4%	-1.21 [-1.90, -0.52]
Gilbody 2015	9.99	6.35	347	9.17	6.34	179	12.1%	0.13 [-0.05, 0.31]
Graaf 2009*	19.75	11.041	190	21.4	11	95	11.6%	-0.15 [-0.40, 0.10]
Kivi 2014	13.23	10.94	30	14.46	9.88	35	9.4%	-0.12 [-0.61, 0.37]
Milgrom 2016	14.5	12.2	21	23	7.5	22	8.0%	-0.83 [-1.45, -0.20]
Mohr 2011	15.43	5.51	40	17	5.68	41	9.9%	-0.28 [-0.72, 0.16]
O´Mahen 2014	11.05	4.71	37	14.26	5.11	34	9.5%	-0.65 [-1.13, -0.17]
Pfeiffer 2020	11.1	4.7	108	11.7	4.1	128	11.5%	-0.14 [-0.39, 0.12]
Watkins 2012	9.36	8.39	33	13	6.25	37	9.5%	-0.49 [-0.97, -0.01]
Total (95% CI)			**931**			**689**	**100.0%**	**-0.44 [-0.73, -0.15]**

Heterogeneity: Tau² = 0.18; Chi² = 62.94, df = 9 (P < 0.00001); I² = 86%
Test for overall effect: Z = 2.94 (P = 0.003)

b) Long-term depression severity

Study or Subgroup	TBI Mean	TBI SD	TBI Total	TAU Mean	TAU SD	TAU Total	Weight	Std. Mean Difference IV, Random, 95% CI
Dennis 2020	6.79	5.4	101	9.77	4.69	96	17.7%	-0.59 [-0.87, -0.30]
Gilbody 2015	7.75	5.82	318	8.45	6.28	166	25.1%	-0.12 [-0.30, 0.07]
Graaf 2009*	16.305	11.07	176	17.5	11.1	91	19.9%	-0.11 [-0.36, 0.15]
Mohr 2011	13.62	6.36	39	14.81	5.01	37	10.0%	-0.21 [-0.66, 0.25]
O´Mahen 2014	8.26	5.5	31	11.14	6.35	29	8.2%	-0.48 [-0.99, 0.03]
Pfeiffer 2020	10.6	5.1	101	11.2	4.4	118	19.0%	-0.13 [-0.39, 0.14]
Total (95% CI)			**766**			**537**	**100.0%**	**-0.24 [-0.41, -0.07]**

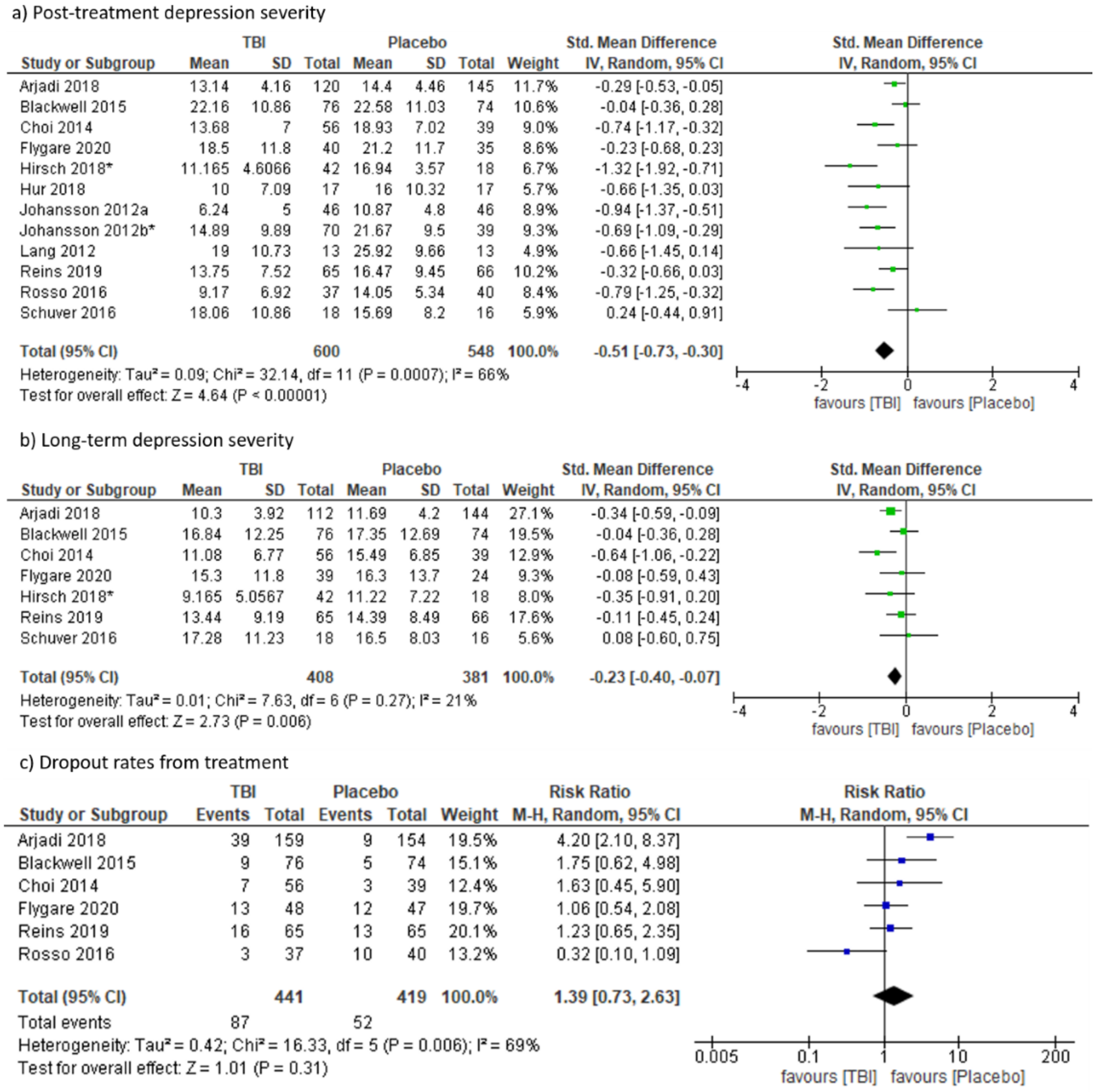
Heterogeneity: Tau² = 0.02; Chi² = 9.67, df = 5 (P = 0.09); I² = 48%
Test for overall effect: Z = 2.79 (P = 0.005)

Note. *multiple treatment arms were summarized.

TBI Versus Attention Placebo

Twelve RCTs tested TBIs against attention placebo controls, which consisted of online psychoeducation [24,37,48,76], participation in an online discussion forum [49], unspecific telephone support calls [32], neutral tasks [42], tasks without training contingency [27,54], symptom monitoring plus short check-in telephone calls [81], daily mood diary [44], and a walking and wellness control condition [83]. Depression severity was significantly lower at posttreatment in the TBI group than in the attention placebo group, with substantial heterogeneity (SMD –0.51, 95% CI –0.73 to –0.30; I^2=66%; 95% PI –1.22 to 0.20). Follow-up depression severity was significantly lower in the TBI group (1 month to 12 months; SMD –0.23, 95% CI –0.40 to –0.07; I^2=21%; 95% PI –0.56 to 0.10). Dropout rates did not differ statistically significantly between groups, with substantial heterogeneity (RR 1.39, 95% CI 0.73 to 2.63; I^2=69; 95% PI 0.56 to 3.43; see Figure 6). Quantitatively exploring heterogeneity for posttreatment depression severity by using prespecified subgroups (technology of intervention delivery, amount of therapist guidance) was not conducted, as the study characteristics were strongly unevenly distributed. It may be possible that heterogeneity was driven by applying broad criteria for attention placebo controls resulting in a rather heterogeneous collection of control conditions. Heterogeneity for dropout rates may be explained by the largest study [24], which clearly favors the attention placebo condition (online psychoeducation) over the TBI condition resulting in low overlap with the other studies in regard to dropout rates. Removing this study from the analysis decreased heterogeneity (I^2=23%) and did not alter the direction of the effect (RR 1.09, 95% CI 0.69 to 1.72). Visual inspection of the funnel plot (Multimedia Appendix 5) was not suspicious.

Figure 6. Forest plots on technology-based psychological intervention versus attention placebo.



a) Post-treatment depression severity

Study or Subgroup	TBI Mean	TBI SD	TBI Total	Placebo Mean	Placebo SD	Placebo Total	Weight	Std. Mean Difference IV, Random, 95% CI
Arjadi 2018	13.14	4.16	120	14.4	4.46	145	11.7%	-0.29 [-0.53, -0.05]
Blackwell 2015	22.16	10.86	76	22.58	11.03	74	10.6%	-0.04 [-0.36, 0.28]
Choi 2014	13.68	7	56	18.93	7.02	39	9.0%	-0.74 [-1.17, -0.32]
Flygare 2020	18.5	11.8	40	21.2	11.7	35	8.6%	-0.23 [-0.68, 0.23]
Hirsch 2018*	11.165	4.6066	42	16.94	3.57	18	6.7%	-1.32 [-1.92, -0.71]
Hur 2018	10	7.09	17	16	10.32	17	5.7%	-0.66 [-1.35, 0.03]
Johansson 2012a	6.24	5	46	10.87	4.8	46	8.9%	-0.94 [-1.37, -0.51]
Johansson 2012b*	14.89	9.89	70	21.67	9.5	39	9.3%	-0.69 [-1.09, -0.29]
Lang 2012	19	10.73	13	25.92	9.66	13	4.9%	-0.66 [-1.45, 0.14]
Reins 2019	13.75	7.52	65	16.47	9.45	66	10.2%	-0.32 [-0.66, 0.03]
Rosso 2016	9.17	6.92	37	14.05	5.34	40	8.4%	-0.79 [-1.25, -0.32]
Schuver 2016	18.06	10.86	18	15.69	8.2	16	5.9%	0.24 [-0.44, 0.91]
Total (95% CI)			**600**			**548**	**100.0%**	**-0.51 [-0.73, -0.30]**

Heterogeneity: Tau² = 0.09; Chi² = 32.14, df = 11 (P = 0.0007); I² = 66%

Test for overall effect: Z = 4.64 (P < 0.00001)

b) Long-term depression severity

Study or Subgroup	TBI Mean	TBI SD	TBI Total	Placebo Mean	Placebo SD	Placebo Total	Weight	Std. Mean Difference IV, Random, 95% CI
Arjadi 2018	10.3	3.92	112	11.69	4.2	144	27.1%	-0.34 [-0.59, -0.09]
Blackwell 2015	16.84	12.25	76	17.35	12.69	74	19.5%	-0.04 [-0.36, 0.28]
Choi 2014	11.08	6.77	56	15.49	6.85	39	12.9%	-0.64 [-1.06, -0.22]
Flygare 2020	15.3	11.8	39	16.3	13.7	24	9.3%	-0.08 [-0.59, 0.43]
Hirsch 2018*	9.165	5.0567	42	11.22	7.22	18	8.0%	-0.35 [-0.91, 0.20]
Reins 2019	13.44	9.19	65	14.39	8.49	66	17.6%	-0.11 [-0.45, 0.24]
Schuver 2016	17.28	11.23	18	16.5	8.03	16	5.6%	0.08 [-0.60, 0.75]
Total (95% CI)			**408**			**381**	**100.0%**	**-0.23 [-0.40, -0.07]**

Heterogeneity: Tau² = 0.01; Chi² = 7.63, df = 6 (P = 0.27); I² = 21%

Test for overall effect: Z = 2.73 (P = 0.006)

c) Dropout rates from treatment

Study or Subgroup	TBI Events	TBI Total	Placebo Events	Placebo Total	Weight	Risk Ratio M-H, Random, 95% CI
Arjadi 2018	39	159	9	154	19.5%	4.20 [2.10, 8.37]
Blackwell 2015	9	76	5	74	15.1%	1.75 [0.62, 4.98]
Choi 2014	7	56	3	39	12.4%	1.63 [0.45, 5.90]
Flygare 2020	13	48	12	47	19.7%	1.06 [0.54, 2.08]
Reins 2019	16	65	13	65	20.1%	1.23 [0.65, 2.35]
Rosso 2016	3	37	10	40	13.2%	0.32 [0.10, 1.09]
Total (95% CI)		**441**		**419**	**100.0%**	**1.39 [0.73, 2.63]**
Total events	87		52			

Heterogeneity: Tau² = 0.42; Chi² = 16.33, df = 5 (P = 0.006); I² = 69%

Test for overall effect: Z = 1.01 (P = 0.31)

Note. *multiple treatment arms were summarized.

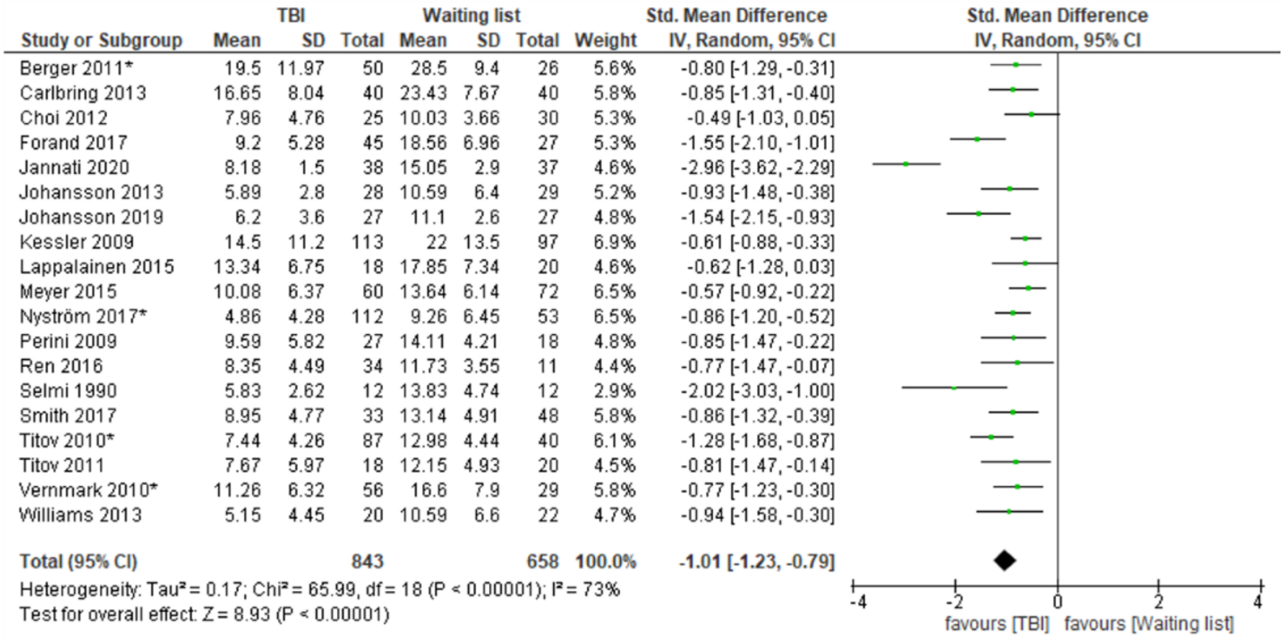
TBI Versus Waitlist Controls

Twenty RCTs tested TBIs against waitlist controls. TBI arms of included studies applied guided [25,29,31,38,46,47,55,65,70,73,84,85,88,89,91], unguided [25,45,62,65,77,95], or therapist-administered [50,91] interventions. All but one study, which examined an offline computer program [84], used internet-based treatment. Depression severity was significantly lower at posttreatment in the TBI group compared to waitlist controls, with substantial heterogeneity (SMD –1.01, 95% CI –1.23 to –0.79; I^2=73%; 95% PI –1.91 to –0.11). Follow-up depression severity was significantly lower in the TBI group, with considerable heterogeneity (2 months to 8 months; SMD –0.74, 95% CI –1.31 to –0.18; I^2=79%; 95% PI –7.24 to 5.76). Dropout rates did not differ between groups (RR 1.13, 95% CI 0.66 to 1.92; I^2=0%; 95% PI 0.04 to 35.12; see Figure 7). Heterogeneity for posttreatment depression severity (I^2=73) may be explained by a potential outlying study [45], which was identified in the course of the search update yielding the largest effect in favor of TBIs (SMD –2.96, 95% CI –3.62 to –2.29) for this comparison. Excluding this study resulted in decreased heterogeneity (I^2=41%) and did not alter the direction of the effect (SMD –0.89, 95% CI –1.04 to –0.74). Heterogeneity for long-term depression severity (I^2=79) may be explained by an older study from 1990 [84], which had a shorter long-term time period (2 months) compared to the other studies (providing 6-month and 8-month long-term data [50,62]). Excluding this study resulted in decreased heterogeneity (I^2=0%) and did not alter the direction of the effect (SMD –0.47, 95% CI –0.70 to

–0.25). The funnel plot (Multimedia Appendix 5) was asymmetrical in the visual inspection.

Figure 7. Forest plots on technology-based psychological intervention versus waitlist.

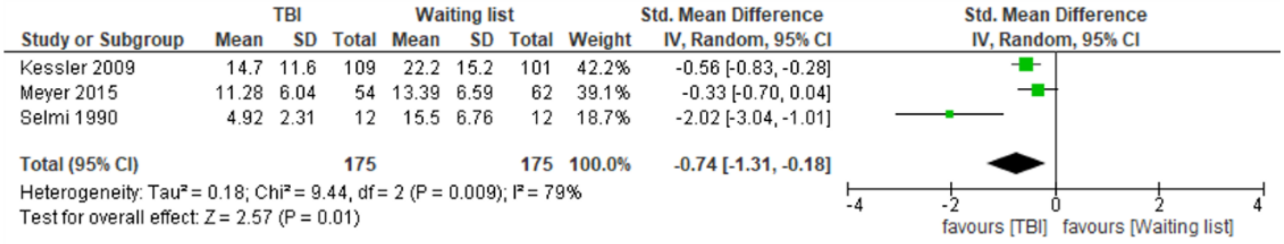
a) Post-treatment depression severity



Study or Subgroup	TBI Mean	TBI SD	TBI Total	Waiting list Mean	Waiting list SD	Waiting list Total	Weight	Std. Mean Difference IV, Random, 95% CI
Berger 2011*	19.5	11.97	50	28.5	9.4	26	5.6%	-0.80 [-1.29, -0.31]
Carlbring 2013	16.65	8.04	40	23.43	7.67	40	5.8%	-0.85 [-1.31, -0.40]
Choi 2012	7.96	4.76	25	10.03	3.66	30	5.3%	-0.49 [-1.03, 0.05]
Forand 2017	9.2	5.28	45	18.56	6.96	27	5.3%	-1.55 [-2.10, -1.01]
Jannati 2020	8.18	1.5	38	15.05	2.9	37	4.6%	-2.96 [-3.62, -2.29]
Johansson 2013	5.89	2.8	28	10.59	6.4	29	5.2%	-0.93 [-1.48, -0.38]
Johansson 2019	6.2	3.6	27	11.1	2.6	27	4.8%	-1.54 [-2.15, -0.93]
Kessler 2009	14.5	11.2	113	22	13.5	97	6.9%	-0.61 [-0.88, -0.33]
Lappalainen 2015	13.34	6.75	18	17.85	7.34	20	4.6%	-0.62 [-1.28, 0.03]
Meyer 2015	10.08	6.37	60	13.64	6.14	72	6.5%	-0.57 [-0.92, -0.22]
Nyström 2017*	4.86	4.28	112	9.26	6.45	53	6.5%	-0.86 [-1.20, -0.52]
Perini 2009	9.59	5.82	27	14.11	4.21	18	4.8%	-0.85 [-1.47, -0.22]
Ren 2016	8.35	4.49	34	11.73	3.55	11	4.4%	-0.77 [-1.47, -0.07]
Selmi 1990	5.83	2.62	12	13.83	4.74	12	2.9%	-2.02 [-3.03, -1.00]
Smith 2017	8.95	4.77	33	13.14	4.91	48	5.8%	-0.86 [-1.32, -0.39]
Titov 2010*	7.44	4.26	87	12.98	4.44	40	6.1%	-1.28 [-1.68, -0.87]
Titov 2011	7.67	5.97	18	12.15	4.93	20	4.5%	-0.81 [-1.47, -0.14]
Vernmark 2010*	11.26	6.32	56	16.6	7.9	29	5.8%	-0.77 [-1.23, -0.30]
Williams 2013	5.15	4.45	20	10.59	6.6	22	4.7%	-0.94 [-1.58, -0.30]
Total (95% CI)			**843**			**658**	**100.0%**	**-1.01 [-1.23, -0.79]**

Heterogeneity: Tau² = 0.17; Chi² = 65.99, df = 18 (P < 0.00001); I² = 73%
Test for overall effect: Z = 8.93 (P < 0.00001)

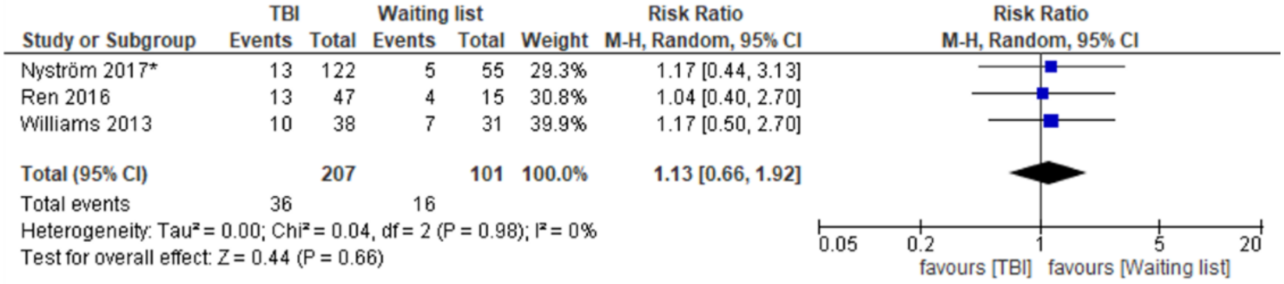
b) Long-term depression severity



Study or Subgroup	TBI Mean	TBI SD	TBI Total	Waiting list Mean	Waiting list SD	Waiting list Total	Weight	Std. Mean Difference IV, Random, 95% CI
Kessler 2009	14.7	11.6	109	22.2	15.2	101	42.2%	-0.56 [-0.83, -0.28]
Meyer 2015	11.28	6.04	54	13.39	6.59	62	39.1%	-0.33 [-0.70, 0.04]
Selmi 1990	4.92	2.31	12	15.5	6.76	12	18.7%	-2.02 [-3.04, -1.01]
Total (95% CI)			**175**			**175**	**100.0%**	**-0.74 [-1.31, -0.18]**

Heterogeneity: Tau² = 0.18; Chi² = 9.44, df = 2 (P = 0.009); I² = 79%
Test for overall effect: Z = 2.57 (P = 0.01)

c) Dropout rates from treatment



Study or Subgroup	TBI Events	TBI Total	Waiting list Events	Waiting list Total	Weight	Risk Ratio M-H, Random, 95% CI
Nyström 2017*	13	122	5	55	29.3%	1.17 [0.44, 3.13]
Ren 2016	13	47	4	15	30.8%	1.04 [0.40, 2.70]
Williams 2013	10	38	7	31	39.9%	1.17 [0.50, 2.70]
Total (95% CI)		**207**		**101**	**100.0%**	**1.13 [0.66, 1.92]**
Total events	36		16			

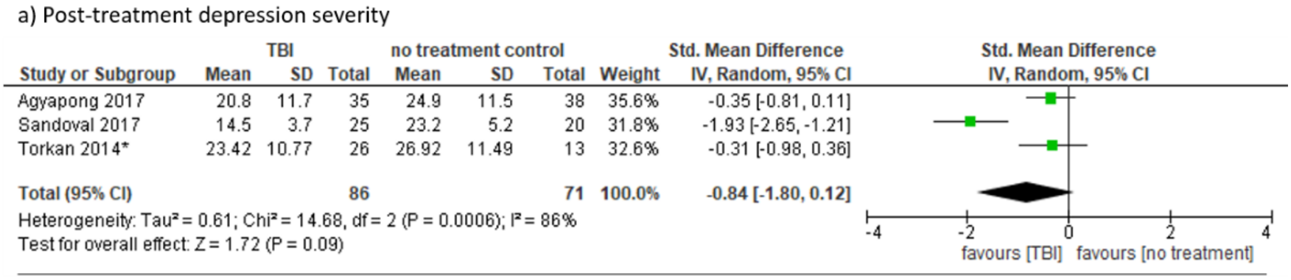
Heterogeneity: Tau² = 0.00; Chi² = 0.04, df = 2 (P = 0.98); I² = 0%
Test for overall effect: Z = 0.44 (P = 0.66)

Note. *multiple treatment arms were summarized.

TBI Versus No-Treatment Control

Three RCTs tested unguided TBIs against no-treatment controls [22,82,90], defined as a comparator where study participants did not receive any offer or encouragement for making use of immediate (eg, TAU) or delayed (eg, waitlist) treatment possibilities. There was no significant difference between TBIs and no-treatment controls at posttreatment (SMD –0.84, 95% CI –1.80 to 0.12; I^2=86%; 95% PI –12.55 to 10.87; see Figure 8). Data on dropout rates were only available for one study [22], indicating that dropout rates did not statistically differ between conditions. Long-term data were not reported. Heterogeneity (I^2=86) may be explained by an outlying, small-sample study with a large CI [82] favoring the TBI condition clearly, which might have been due to the provision of a more intensive TBI, as the TBI is either longer or needs a more active user engagement when compared to the other trials’ interventions. Excluding this study resulted in decreased heterogeneity (SMD –0.34, 95% CI –0.72 to 0.04; I^2=0%) and did not change the direction of the effect.

Figure 8. Forest plot for technology-based psychological intervention versus no-treatment control.



Note. *multiple treatment arms were summarized.

Comparing Different Types of TBIs

Overall, 21 studies compared different TBIs competitively, 12 of which [25,34,40,42,49,65,68,70,80,88,90,91] compared multiple (2 or more) TBIs with a control group (eg, TAU). Thus, certain arms of these studies were suitable for other prespecified comparisons (eg, Gilbody et al [40] for TBI vs TAU). Nine of them compared TBIs versus another TBI [30,33,56,60,75,86,93,96,98] without having a further control group. For these studies, meta-analysis was not computed, since research foci of studies were too heterogeneous—they investigated different types of guidance (eg, telephone support vs email support) [56,75,98], treatment approaches [30,33,60,86,96], or delivery modes [93].

Other Comparisons

Two studies were identified during the search update that could not be matched to our comparisons [71,94]. One study compared a guided web-based CBT tool (iFightDepression) against an active control intervention receiving progressive muscle relaxation provided via a download link [71]. Another study investigated a TBI in combination with and without transcranial direct current stimulation [94].

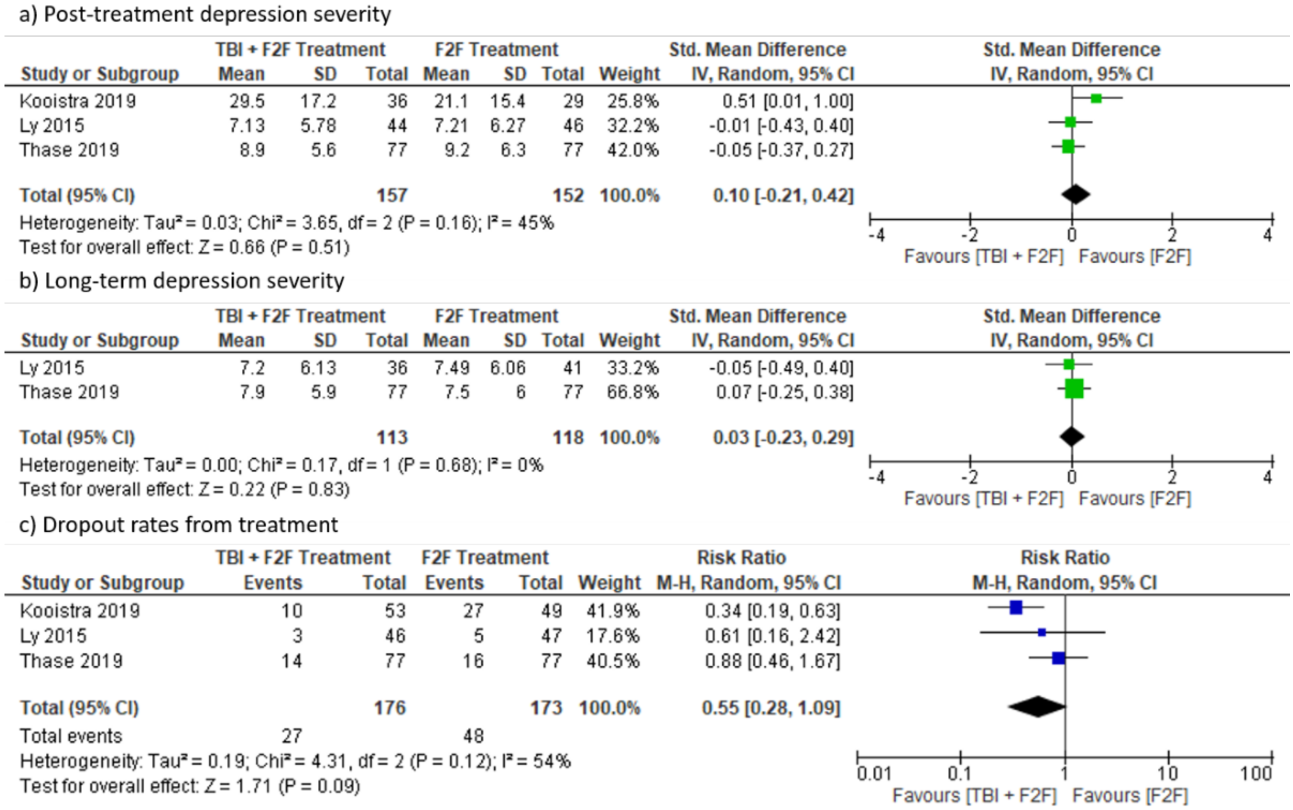
Blended Treatments

11 RCTs tested blended treatments against different f2f treatments. Six RCTs were identified combining TBIs with f2f psychotherapy versus f2f psychotherapy alone. In these trials, TBIs were delivered in addition to outpatient psychotherapy [26,52,97], inpatient psychotherapy [99], and psychotherapy treatment sessions where the setting was not specified [59,87]. Two RCTs were identified comparing a TBI in addition to medication versus medication alone [53,61], and 2 RCTs tested a TBI with f2f TAU against TAU [28,68]. Additionally, we identified one RCT [69] where blended treatment (f2f CBT and internet-based CBT) was provided alongside TAU (psychiatric treatment) compared to TAU. Overall, 8 superiority [26,28,53,61,68,69,97,99] and 3 noninferiority trials [52,59,87] applying blended treatments were identified.

Noninferiority Trials

There was no statistically significant difference between groups concerning posttreatment depression severity (SMD 0.10, 95% CI –0.21 to 0.42; I^2=45%; 95% PI –2.91 to 3.12), long-term (6 months) depression severity (SMD 0.03, 95% CI –0.23 to 0.29; I^2=0%), or dropouts (RR 0.55, 95% CI 0.28 to 1.09; I^2=54%; 95% PI 0 to 663.21; see Figure 9).

Figure 9. Forest plots for blended treatments (noninferiority trials).



a) Post-treatment depression severity

Study or Subgroup	TBI + F2F Treatment Mean	SD	Total	F2F Treatment Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Kooistra 2019	29.5	17.2	36	21.1	15.4	29	25.8%	0.51 [0.01, 1.00]
Ly 2015	7.13	5.78	44	7.21	6.27	46	32.2%	-0.01 [-0.43, 0.40]
Thase 2019	8.9	5.6	77	9.2	6.3	77	42.0%	-0.05 [-0.37, 0.27]
Total (95% CI)			**157**			**152**	**100.0%**	**0.10 [-0.21, 0.42]**

Heterogeneity: Tau² = 0.03; Chi² = 3.65, df = 2 (P = 0.16); I² = 45%
Test for overall effect: Z = 0.66 (P = 0.51)

b) Long-term depression severity

Study or Subgroup	TBI + F2F Treatment Mean	SD	Total	F2F Treatment Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Ly 2015	7.2	6.13	36	7.49	6.06	41	33.2%	-0.05 [-0.49, 0.40]
Thase 2019	7.9	5.9	77	7.5	6	77	66.8%	0.07 [-0.25, 0.38]
Total (95% CI)			**113**			**118**	**100.0%**	**0.03 [-0.23, 0.29]**

Heterogeneity: Tau² = 0.00; Chi² = 0.17, df = 1 (P = 0.68); I² = 0%
Test for overall effect: Z = 0.22 (P = 0.83)

c) Dropout rates from treatment

Study or Subgroup	TBI + F2F Treatment Events	Total	F2F Treatment Events	Total	Weight	Risk Ratio M-H, Random, 95% CI
Kooistra 2019	10	53	27	49	41.9%	0.34 [0.19, 0.63]
Ly 2015	3	46	5	47	17.6%	0.61 [0.16, 2.42]
Thase 2019	14	77	16	77	40.5%	0.88 [0.46, 1.67]
Total (95% CI)		**176**		**173**	**100.0%**	**0.55 [0.28, 1.09]**
Total events	27		48			

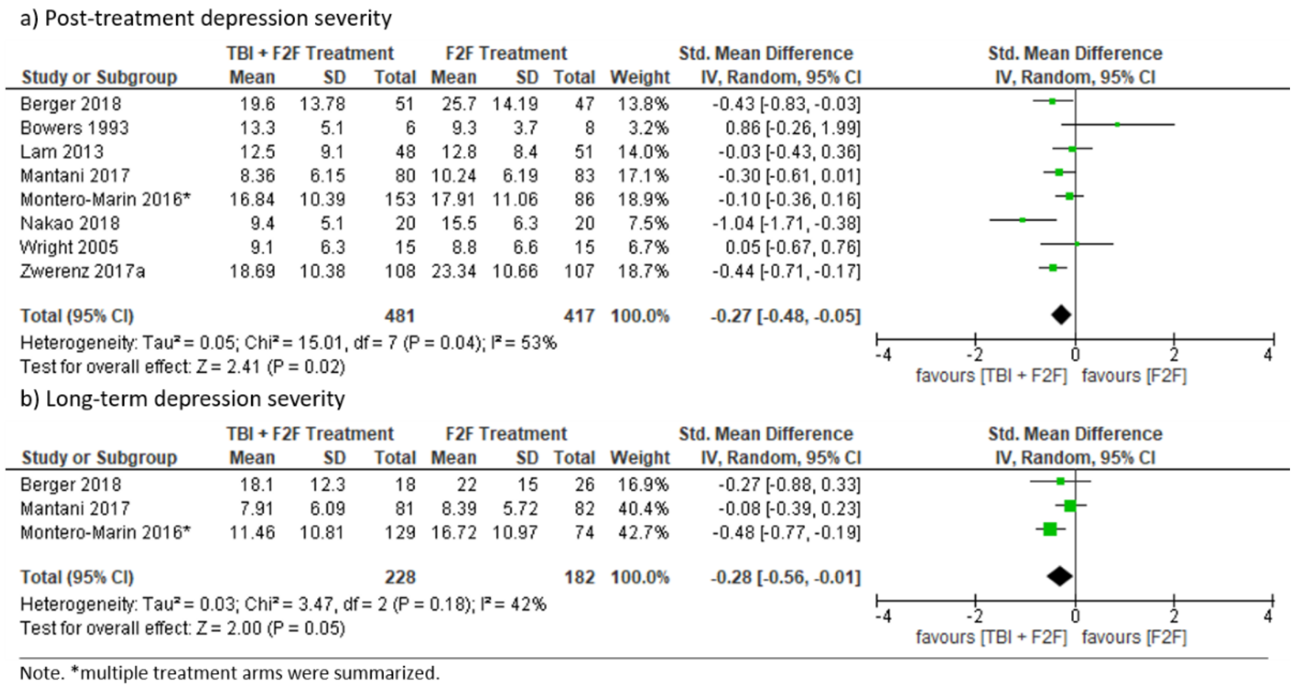
Heterogeneity: Tau² = 0.19; Chi² = 4.31, df = 2 (P = 0.12); I² = 54%
Test for overall effect: Z = 1.71 (P = 0.09)

Superiority Trials

Depression severity was significantly lower at posttreatment in blended treatment groups compared to f2f treatment controls, with substantial heterogeneity (SMD –0.27, 95% CI –0.48 to –0.05; I^2=53%; 95% PI –0.88 to 0.34). Treatments did not differ significantly concerning long-term (4 months to 15 months) depression severity (SMD –0.28, 95% CI –0.56 to –0.01; I^2=42%; 95% PI –3.13 to 2.57). There were no data available for dropouts concerning superiority trials (see Figure 10).

Heterogeneity (I^2=53%) for posttreatment depression severity may be explained by an outlying, small-sample study [69] favoring the blended treatment condition more clearly, which might have been due to the provision of a more intensive treatment regimen, since patients received blended treatment (internet-based TBI combined with f2f CBT) in addition to TAU consisting of f2f psychiatric care. Excluding this study resulted in decreased heterogeneity (SMD –0.22, 95% CI –0.40 to –0.03; I^2=37%) and did not change the direction of effect.

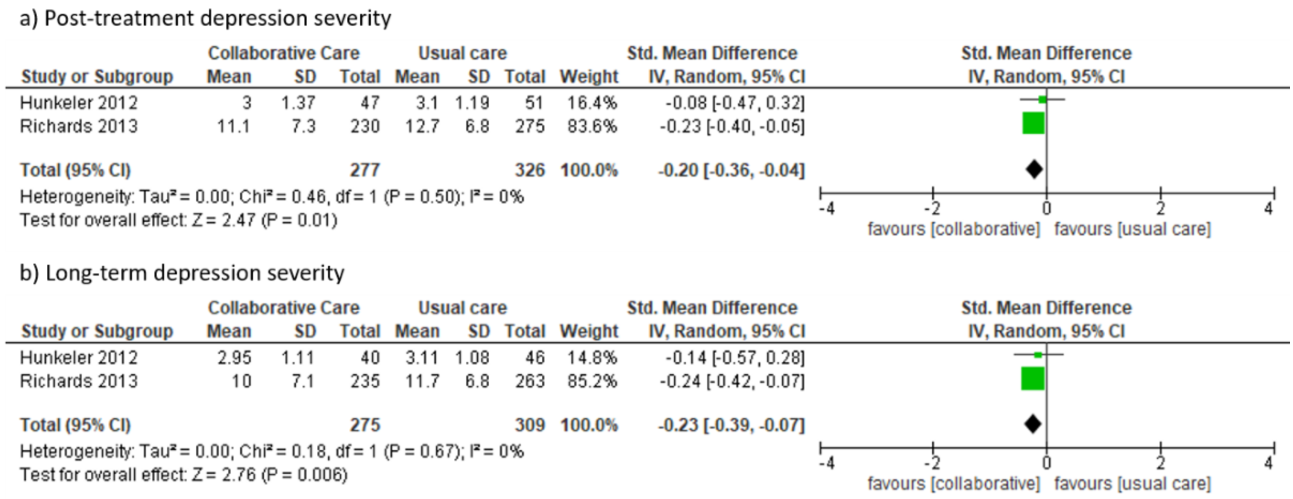
Figure 10. Forest plots for blended treatments (superiority trials).



Collaborative Care Approach

Three RCTs were identified applying TBIs, which were tested against usual care arms [43,79,80], in the context of a collaborative care approach. TBIs delivered in the context of collaborative care trials yielded lower posttreatment (SMD –0.20, 95% CI –0.36 to –0.04; I^2=0%) and long-term (12 months: SMD –0.23, 95% CI –0.39 to –0.07; I^2=0%) depression severity compared to usual care arms (see Figure 11).

Figure 11. Forest plots for collaborative care approaches.



Stepped Care Approach

Two RCTs using TBIs in the context of a stepped care approach were identified in the course of the search update. The studies were too heterogeneous for evidence syntheses, since one study tested a stepped care approach (first step: internet-based treatment, second step: telephone-based treatment) against telephone-based treatment alone [67], and the other study tested an internet-based intervention against a waitlist control group as a first step within a stepped care approach [78].

Discussion

Principal Findings

Our study found that when compared to different control conditions, TBIs were more effective not only when delivered as stand-alone interventions but also when they were delivered as blended treatments or in collaborative care trials for people with diagnosed depression. Dropout rates did not differ between TBI and control conditions; however, assessment of TBI acceptance was limited due to underpowered comparisons. In addition, relevant statistical heterogeneity was a common finding for most meta-analytical comparisons. We included 78 RCTs comprising different application formats (stand-alone interventions [61/78, 78%], blended treatments [12/78, 15%], and stepped care [2/78, 3%] or collaborative care trials [3/78, 4%]), interventions, technologies for intervention delivery, clinical populations, and control groups.

Stand-Alone Interventions

TBIs showed comparable effects to f2f treatments. Our findings are in line with a previous meta-analyses that found equivalent overall effects when comparing internet-based CBT to f2f treatment for mental disorders and somatic conditions on posttreatment symptom burden for studies on depressive symptoms specifically and for dropouts rates [100]. However, both results should be interpreted with caution, since both evidence syntheses were based on a limited number of studies.

When TBIs were tested against TAU controls, we found medium-to-small effects favoring TBIs concerning posttreatment and long-term depression severity. TAU was heterogeneous and consisted mostly of a mix of treatment options depending on the resources and routines of health care providers, general practitioner care, or care delivered in outpatient clinics. In addition, two-thirds of the studies included for this comparison also provided TAU in the TBI condition. Our results are in line with 2 previous meta-analyses that found a small effect favoring TBIs in comparison with TAU [101,102].

TBIs yielded beneficial medium-to-small effects on posttreatment and long-term depression severity when compared to attention placebo controls. To our knowledge, there is no previous meta-analysis available on this issue. However, the results are comparable to those comparing f2f psychotherapy with placebo [103] and pill placebo control groups [104].

We found a large effect in favor of the TBI group compared to waitlist controls for posttreatment and long-term depression severity. Our findings are in line with the only existing meta-analysis investigating TBIs in people with diagnosed depression [13]. This is not surprising, as there was a high overlap between the included studies. However, we were able to include more RCTs (+10) for the comparison of TBIs versus waitlist controls due to broader inclusion criteria and an updated literature search. Thus, our analysis emphasizes the robustness of the previous findings. However, the funnel plot on posttreatment depression severity was asymmetrical, with an emphasis on small studies depicting large differences in favor of TBIs compared to waitlist controls. Nevertheless, this is not a clear indicator of reporting bias because there are other sources (eg, heterogeneity, poor methodological quality) causing funnel plot asymmetry [16]. Between-study heterogeneity seems plausible to partly explain asymmetry, since we applied broad eligibility criteria and suspicious studies differed from the others in terms of population (postpartum depression) or publication year (1990), potentially resulting in more elevated differences.

Finally, TBIs did not result in lower posttreatment depression severity scores than no-treatment controls. This was not reasonable to expect, since no-treatment controls are comparable weak control groups, such as waitlist controls, which yield large effects when compared to TBIs [13]. Moreover, based on study reports, it cannot be ruled out that people allocated to the no-treatment control group made use of other health services for depression complaints (eg, care by a general practitioner), thus questioning whether true no-treatment controls were applied.

Blended Treatments

We identified a small effect favoring blended treatments delivered in a superiority trial design compared to f2f treatments concerning posttreatment depression severity. Meta-analysis on blended treatments delivered in a noninferiority trial design (ie, substantial shortening of f2f contacts) did not reveal differences in posttreatment or long-term depression severity or on dropout rates compared to f2f treatments. To the best of our knowledge, there is no previous meta-analysis investigating the effectiveness and acceptance of blended treatments in people with depression. Additionally, despite extensive discussions on their potential usefulness for mental health care [105,106], there is no uniform definition of blended care/treatment as they are operationalized in different ways and rationales for blended treatments are often missing [105]. This was also the case in our study, since the concept of combining a TBI with an f2f treatment was usually explained insufficiently or not at all. In the included studies, it appears that blended treatments were implemented based on the motto more is more (intensification of the therapeutic dose by providing add-on treatment following a superiority trial design). Nevertheless, future studies could define and investigate more sophisticated variants of blended treatments, since there are many useful possibilities to enrich onsite therapy by, for example, fostering preprocessing and postprocessing of sessions or for diagnostic purposes in everyday life (eg, self-monitoring) [106].

Collaborative Care Approach

TBIs delivered in the context of collaborative care yielded small effects on posttreatment and long-term depression severity when compared to TAU controls. However, findings should be viewed with caution, since only a few studies have been available until now, and investigated collaborative care approaches are heterogeneous. The identified posttreatment and long-term effects on depression severity are comparable to reported effects investigating collaborative care approaches without TBIs in comparison to usual care [107,108]. However, we do not know if and how much the technology-based component is involved in the effectiveness of these interventions, since collaborative care approaches are complex. Testing collaborative care approaches with and without a TBI component may help to determine the add-on benefit of this element and may be concurrently useful for a comparative cost-benefit analysis.

Strengths and Limitations

Our review was conducted in line with Cochrane standards [16] and reported following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [109]. Additionally, studies were selected according to prespecified criteria [15]. We conducted a highly sensitive literature search considering key databases, databases of grey literature, and clinical trial registries without limiting the literature search to language. However, because of the extensive literature search, we deviated from the study protocol by omitting the forward and backward reference search. We structured and synthesized evidence using prespecified comparisons defined in the study protocol covering different application formats of TBIs in the acute treatment phase.

We applied broad inclusion criteria [15] contributing to observed heterogeneity regarding interventions, technologies for intervention delivery, psychotherapeutic rationales, and clinical populations in the included studies. Unfortunately, we were not able to explain statistical heterogeneity quantitatively (eg, by subgroup analyses) for most comparisons, since often only a few studies were available. However, we tried to explore heterogeneity narratively in these cases. In addition, when heterogeneity of the included studies is present (ie, $I^2 \neq 0$), the CI covers a narrower range than the PI of the respective comparison. Thus, pooled effects (SMDs) should be interpreted with caution: It may be that even if the pooled effect is significant (ie, CI not crossing null), the corresponding PI covers the null effect, meaning that in a new study conducted in a different setting (eg, different population), null treatment effects or effects in the other direction (harmful) may occur [20,110].

Although some information on dropouts [11] or treatment adherence [111] is addressed by most RCTs in this field, a comprehensive assessment of TBI acceptance was only partially possible, since data on dropouts were either missing or not usable (eg, data were only provided for one arm) or meta-analytic calculations were not possible (when no dropouts occurred in both study arms).

Considering the risk of bias ratings when interpreting the results, we found that the most common source of risk of bias was nonblinding of participants and personnel, followed by selective reporting and other bias. However, blinding of study participants is rarely possible in trials on TBIs.

Conclusions

TBIs delivered as stand-alone interventions, blended treatments, or in collaborative care trials yield mostly beneficial effects in people with diagnosed depression. By investigating different application formats of TBIs, people being diagnosed with depression, and the long-term effectiveness of interventions, our results may be especially helpful to inform routine care. Given the potential transferability of our findings to routine care, we think that our findings may represent effectiveness (effectiveness under routine care), rather than efficacy (effectiveness under ideal conditions) of findings. Additionally, our results show a very consistent image of TBIs (it works), despite the clinical and methodological heterogeneity of the included studies.

However, there are still open questions that need to be addressed in future research. Even though dropouts are by far the most reported indicator for treatment acceptance/patient safety in studies with TBIs [11], data were often not usable for data synthesis resulting in underpowered comparisons for safety/acceptance assessment. Therefore, our findings with regard to this outcome should be interpreted with caution.

Additionally, safety assessments of TBIs considering different types of safety measures in people with diagnosed depression have not yet been conducted. Thus, to obtain a more comprehensive impression of the safety of TBIs, we suggest including all indicators according to Rozenthal et al [112] to evaluate negative events: (severe) adverse events, dropouts, nonresponse, novel symptoms, and unwanted events.

Acknowledgments

We thank Eileen Wehmann, Mareike Dreier, and Tharanya Seeralan for their support in extracting data. This study was funded by grant 01KG1705 (funding period: 08/2017 to 10/2019) from the Federal Ministry of Education and Research (in German, Bundesministerium für Bildung und Forschung). The funding institution had no role in the design of this study; its execution, analyses, interpretation of the data; or the decision to submit results.

Conflicts of Interest

MK is a psychotherapist in training (cognitive behavioral therapy). MH and SL are licensed psychotherapists (cognitive behavioral therapy). SL is additionally employed at the Institute for Psychotherapy at the University Medical Center Hamburg-Eppendorf, which provides psychotherapist training. MH and HB are participating in the current revision of the German S3 national clinical practice guideline on the treatment of adults with unipolar depression. HB received consultancy fees, reimbursement of congress attendance and travel costs, and payments for lectures from psychotherapy and psychiatry associations as well as psychotherapy training institutes in the context of e-mental health topics. He has been the beneficiary of e-mental health study support (third-party funding) from several public funding organizations. LK declares that he has no competing interests.

Multimedia Appendix 1
Summarizing table for meta-analysis.
[PDF File (Adobe PDF File), 119 KB - jmir_v23i6e24584_app1.pdf]

Multimedia Appendix 2
Characteristics of included studies.
[PDF File (Adobe PDF File), 279 KB - jmir_v23i6e24584_app2.pdf]

Multimedia Appendix 3
Therapeutic rationale for technology-based psychological interventions.
[PDF File (Adobe PDF File), 125 KB - jmir_v23i6e24584_app3.pdf]

Multimedia Appendix 4
Risk of bias ratings (study level).
[PDF File (Adobe PDF File), 150 KB - jmir_v23i6e24584_app4.pdf]

Multimedia Appendix 5
Funnel plots.
[PDF File (Adobe PDF File), 79 KB - jmir_v23i6e24584_app5.pdf]

References

1. DGPPN, BÄK, KBV, AWMF for the guideline group unipolar depression. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression – Langfassung, 2. Auflage. Version 5. 2015. URL: https://www.leitlinien.de/themen/depression/pdf/depression-2aufl-vers5-lang.pdf
2. Cabello M, Mellor-Marsá B, Sabariego C, Cieza A, Bickenbach J, Ayuso-Mateos JL. Psychosocial features of depression: a systematic literature review. J Affect Disord 2012 Dec 01;141(1):22-33. [doi: 10.1016/j.jad.2011.12.009] [Medline: 22209189]
3. World Health Organization. The global burden of disease: 2004 update. Geneva: World Health Organization; 2008. URL: https://apps.who.int/iris/bitstream/handle/10665/43942/9789241563710_eng.pdf [accessed 2018-10-04]
4. National Institute for Health and Care Excellence. Depression: the treatment and management of depression in adults: full guideline (draft for consultation). National Institute for Health and Care Excellence. 2017. URL: https://www.nice.org.uk/guidance/gid-cgwave0725/documents/draft-guideline [accessed 2018-05-02]
5. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
6. Köhnen M, Dirmaier J, Härter M. [Potentials and Challenges of E-Mental Health Interventions in Mental Health Care]. Fortschr Neurol Psychiatr 2019 Mar;87(3):160-164. [doi: 10.1055/a-0853-2568] [Medline: 30891717]
7. 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]
8. 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]
9. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
10. Huguet A, Miller A, Kisely S, Rao S, Saadat N, McGrath PJ. A systematic review and meta-analysis on the efficacy of Internet-delivered behavioral activation. J Affect Disord 2018 Dec 01;235:27-38. [doi: 10.1016/j.jad.2018.02.073] [Medline: 29649708]
11. Köhnen M, Dreier M, Seeralan T, Kriston L, Härter M, Baumeister H, et al. Evidence on technology-based psychological interventions in diagnosed depression: systematic review. JMIR Ment Health 2021 Feb 10;8(2):e21700 [FREE Full text] [doi: 10.2196/21700] [Medline: 33565981]
12. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance Priority Setting Partnership. Lancet Psychiatry 2018 Oct;5(10):845-854. [doi: 10.1016/S2215-0366(18)30296-7] [Medline: 30170964]
13. Königbauer J, Letsch J, Doebler P, Ebert D, Baumeister H. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
14. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
15. Köhnen M, Kriston L, Härter M, Dirmaier J, Liebherz S. Rationale and design of a systematic review: effectiveness and acceptance of technology-based psychological interventions in different clinical phases of depression management. BMJ Open 2019 Mar 27;9(3):e028042 [FREE Full text] [doi: 10.1136/bmjopen-2018-028042] [Medline: 30918040]
16. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration. 2011 Mar. URL: https://handbook-5-1.cochrane.org/ [accessed 2021-05-21]
17. Machmutow K, Meister R, Jansen A, Kriston L, Watzke B, Härter MC, et al. Comparative effectiveness of continuation and maintenance treatments for persistent depressive disorder in adults. Cochrane Database Syst Rev 2019 May 20;5:CD012855 [FREE Full text] [doi: 10.1002/14651858.CD012855.pub2] [Medline: 31106850]

18. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986 Sep;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
19. Kriston L. Dealing with clinical heterogeneity in meta-analysis: assumptions, methods, interpretation. Int J Methods Psychiatr Res 2013 Mar;22(1):1-15 [FREE Full text] [doi: 10.1002/mpr.1377] [Medline: 23494781]
20. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011 Feb 10;342:d549. [doi: 10.1136/bmj.d549] [Medline: 21310794]
21. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
22. Agyapong VIO, Juhás M, Ohinmaa A, Omeje J, Mrklas K, Suen VYM, et al. Randomized controlled pilot trial of supportive text messages for patients with depression. BMC Psychiatry 2017 Aug 02;17(1):286 [FREE Full text] [doi: 10.1186/s12888-017-1448-2] [Medline: 28768493]
23. Andersson G, Hesser H, Veilord A, Svedling L, Andersson F, Sleman O, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord 2013 Dec;151(3):986-994. [doi: 10.1016/j.jad.2013.08.022] [Medline: 24035673]
24. Arjadi R, Nauta MH, Scholte WF, Hollon SD, Chowdhary N, Suryani AO, et al. Internet-based behavioural activation with lay counsellor support versus online minimal psychoeducation without support for treatment of depression: a randomised controlled trial in Indonesia. Lancet Psychiatry 2018 Sep;5(9):707-716. [doi: 10.1016/s2215-0366(18)30223-2]
25. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011 Dec;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
26. Berger T, Krieger T, Sude K, Meyer B, Maercker A. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: results of a pragmatic randomized controlled trial. J Affect Disord 2018 Feb;227:455-462. [doi: 10.1016/j.jad.2017.11.021] [Medline: 29154168]
27. Blackwell SE, Browning M, Mathews A, Pictet A, Welch J, Davies J, et al. Positive imagery-based cognitive bias modification as a web-based treatment tool for depressed adults: a randomized controlled trial. Clin Psychol Sci 2015 Jan;3(1):91-111 [FREE Full text] [doi: 10.1177/2167702614560746] [Medline: 25984421]
28. Bowers W, Stuart S, Macfarlane R, Gorman L. Use of computer-administered cognitive-behavior therapy with depressed inpatients. Depression 1993;1(6):294-299. [doi: 10.1002/depr.3050010603]
29. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki Å, Vernmark K, et al. Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. J Affect Di 2013 Jun;148(2-3):331-337. [doi: 10.1016/j.jad.2012.12.020] [Medline: 23357657]
30. Celano CM, Beale EE, Mastromauro CA, Stewart JG, Millstein RA, Auerbach RP, et al. Psychological interventions to reduce suicidality in high-risk patients with major depression: a randomized controlled trial. Psychol Med 2017 Apr;47(5):810-821 [FREE Full text] [doi: 10.1017/S0033291716002798] [Medline: 27876105]
31. Choi I, Zou J, Titov N, Dear BF, Li S, Johnston L, et al. Culturally attuned Internet treatment for depression amongst Chinese Australians: a randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]
32. Choi NG, Marti CN, Bruce ML, Hegel MT, Wilson NL, Kunik ME. Six-month postintervention depression and disability outcomes of in-home telehealth problem-solving therapy for depressed, low-income homebound older adults. Depress Anxiety 2014 Aug;31(8):653-661 [FREE Full text] [doi: 10.1002/da.22242] [Medline: 24501015]
33. Corruble E, Swartz HA, Bottai T, Vaiva G, Bayle F, Llorca P, et al. Telephone-administered psychotherapy in combination with antidepressant medication for the acute treatment of major depressive disorder. J Affect Disord 2016 Jan 15;190:6-11. [doi: 10.1016/j.jad.2015.07.052] [Medline: 26480205]
34. de Graaf L, Gerhards S, Arntz A, Riper H, Metsemakers J, Evers S, et al. Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. Br J Psychiatry 2009 Jul;195(1):73-80. [doi: 10.1192/bjp.bp.108.054429] [Medline: 19567900]
35. Dennis C, Grigoriadis S, Zupancic J, Kiss A, Ravitz P. Telephone-based nurse-delivered interpersonal psychotherapy for postpartum depression: nationwide randomised controlled trial. Br J Psychiatry 2020 Apr;216(4):189-196. [doi: 10.1192/bjp.2019.275] [Medline: 32029010]
36. Egede LE, Acierno R, Knapp RG, Lejuez C, Hernandez-Tejada M, Payne EH, et al. Psychotherapy for depression in older veterans via telemedicine: a randomised, open-label, non-inferiority trial. Lancet Psychiatry 2015 Aug;2(8):693-701. [doi: 10.1016/S2215-0366(15)00122-4] [Medline: 26249300]
37. Flygare A, Engström I, Hasselgren M, Jansson-Fröjmark M, Frejgrim R, Andersson G, et al. Internet-based CBT for patients with depressive disorders in primary and psychiatric care: is it effective and does comorbidity affect outcome? Internet Interv 2020 Mar;19:100303 [FREE Full text] [doi: 10.1016/j.invent.2019.100303] [Medline: 32055451]
38. Forand NR, Barnett JG, Strunk DR, Hindiyeh MU, Feinberg JE, Keefe JR. Efficacy of guided iCBT for depression and mediation of change by cognitive skill acquisition. Behav Ther 2018 Dec;49(2):295-307 [FREE Full text] [doi: 10.1016/j.beth.2017.04.004] [Medline: 29530267]

39. Forsell E, Bendix M, Holländare F, Szymanska von Schultz B, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: a randomised controlled trial. J Affect Disord 2017 Oct 15;221:56-64 [FREE Full text] [doi: 10.1016/j.jad.2017.06.013] [Medline: 28628768]
40. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
41. Gili M, Castro A, García-Palacios A, Garcia-Campayo J, Mayoral-Cleries F, Botella C, Pérez-Ara, et al. Efficacy of three low-intensity, internet-based psychological interventions for the treatment of depression in primary care: randomized controlled trial. J Med Internet Res 2020 Jun 05;22(6):e15845 [FREE Full text] [doi: 10.2196/15845] [Medline: 32501276]
42. Hirsch C, Krahé C, Whyte J, Loizou S, Bridge L, Norton S, et al. Interpretation training to target repetitive negative thinking in generalized anxiety disorder and depression. J Consult Clin Psychol 2018 Dec;86(12):1017-1030. [doi: 10.1037/ccp0000310] [Medline: 30507227]
43. Hunkeler EM, Hargreaves WA, Fireman B, Terdiman J, Meresman JF, Porterfield Y, et al. A web-delivered care management and patient self-management program for recurrent depression: a randomized trial. Psychiatr Serv 2012 Nov;63(11):1063-1071. [doi: 10.1176/appi.ps.005332011] [Medline: 22983558]
44. Hur J, Kim B, Park D, Choi S. A scenario-based cognitive behavioral therapy mobile app to reduce dysfunctional beliefs in individuals with depression: a randomized controlled trial. Telemed J E Health 2018 Dec;24(9):710-716. [doi: 10.1089/tmj.2017.0214] [Medline: 29323626]
45. Jannati N, Mazhari S, Ahmadian L, Mirzaee M. Effectiveness of an app-based cognitive behavioral therapy program for postpartum depression in primary care: a randomized controlled trial. Int J Med Inform 2020 Sep;141:104145. [doi: 10.1016/j.ijmedinf.2020.104145] [Medline: 32480319]
46. Johansson O, Bjärehed J, Andersson G, Carlbring P, Lundh L. Effectiveness of guided internet-delivered cognitive behavior therapy for depression in routine psychiatry: a randomized controlled trial. Internet Interv 2019 Sep;17:100247 [FREE Full text] [doi: 10.1016/j.invent.2019.100247] [Medline: 31249791]
47. Johansson R, Björklund M, Hornborg C, Karlsson S, Hesser H, Ljótsson B, et al. Affect-focused psychodynamic psychotherapy for depression and anxiety through the Internet: a randomized controlled trial. PeerJ 2013;1:e102 [FREE Full text] [doi: 10.7717/peerj.102] [Medline: 23862104]
48. Johansson R, Ekbladh S, Hebert A, Lindström M, Möller S, Petitt E, et al. Psychodynamic guided self-help for adult depression through the internet: a randomised controlled trial. PLoS One 2012;7(5):e38021 [FREE Full text] [doi: 10.1371/journal.pone.0038021] [Medline: 22741027]
49. Johansson R, Sjöberg E, Sjögren M, Johnsson E, Carlbring P, Andersson T, et al. Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: a randomized controlled trial. PLoS One 2012;7(5):e36905 [FREE Full text] [doi: 10.1371/journal.pone.0036905] [Medline: 22615841]
50. Kessler D, Lewis G, Kaur S, Wiles N, King M, Weich S, et al. Therapist-delivered Internet psychotherapy for depression in primary care: a randomised controlled trial. Lancet 2009 Aug 22;374(9690):628-634. [doi: 10.1016/S0140-6736(09)61257-5] [Medline: 19700005]
51. Kivi M, Eriksson MCM, Hange D, Petersson E, Vernmark K, Johansson B, et al. Internet-based therapy for mild to moderate depression in Swedish primary care: short term results from the PRIM-NET randomized controlled trial. Cogn Behav Ther 2014;43(4):289-298 [FREE Full text] [doi: 10.1080/16506073.2014.921834] [Medline: 24911260]
52. Kooistra LC, Wiersma JE, 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]
53. Lam RW, Parikh SV, Ramasubbu R, Michalak EE, Tam EM, Axler A, et al. Effects of combined pharmacotherapy and psychotherapy for improving work functioning in major depressive disorder. Br J Psychiatry 2013 Nov;203(5):358-365. [doi: 10.1192/bjp.bp.112.125237] [Medline: 24029535]
54. Lang TJ, Blackwell SE, Harmer CJ, Davison P, Holmes EA. Cognitive bias modification using mental imagery for depression: developing a novel computerized intervention to change negative thinking styles. Eur J Pers 2012 Mar;26(2):145-157 [FREE Full text] [doi: 10.1002/per.855] [Medline: 23316101]
55. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-based acceptance and commitment therapy for depressive symptoms with minimal support: a randomized controlled trial. Behav Modif 2015 Aug 6;39(6):805-834. [doi: 10.1177/0145445515598142] [Medline: 26253644]
56. Lindner P, Olsson EL, Johnsson A, Dahlin M, Andersson G, Carlbring P. The impact of telephone versus e-mail therapist guidance on treatment outcomes, therapeutic alliance and treatment engagement in Internet-delivered CBT for depression: a randomised pilot trial. Internet Interv 2014 Oct;1(4):182-187. [doi: 10.1016/j.invent.2014.09.001]
57. Löbner M, Pabst A, Stein J, Dorow M, Matschinger H, Luppa M, et al. Computerized cognitive behavior therapy for patients with mild to moderately severe depression in primary care: a pragmatic cluster randomized controlled trial (@ktiv). J Affect Disord 2018 Oct 01;238:317-326. [doi: 10.1016/j.jad.2018.06.008] [Medline: 29902736]

58. Luxton DD, Pruitt LD, Wagner A, Smolenski DJ, Jenkins-Guarnieri MA, Gahm G. Home-based telebehavioral health for U.S. military personnel and veterans with depression: a randomized controlled trial. J Consult Clin Psychol 2016 Nov;84(11):923-934. [doi: 10.1037/ccp0000135] [Medline: 27599225]
59. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
60. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
61. Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. J Med Internet Res 2017 Nov 03;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
62. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an Internet intervention (Deprexis) on severe depression symptoms: randomized controlled trial. Internet Interv 2015 Mar;2(1):48-59. [doi: 10.1016/j.invent.2014.12.003]
63. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. J Med Internet Res 2016;18(3):e54 [FREE Full text] [doi: 10.2196/jmir.4993] [Medline: 26952645]
64. Mohr DC, Carmody T, Erickson L, Jin L, Leader J. Telephone-administered cognitive behavioral therapy for veterans served by community-based outpatient clinics. J Consult Clin Psychol 2011 Apr;79(2):261-265. [doi: 10.1037/a0022395] [Medline: 21299274]
65. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013 Aug;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
66. Mohr DC, Ho J, Duffecy J, Reifler D, Sokol L, Burns MN, et al. Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients: a randomized trial. JAMA 2012 Jun 06;307(21):2278-2285 [FREE Full text] [doi: 10.1001/jama.2012.5588] [Medline: 22706833]
67. Mohr DC, Lattie EG, Tomasino KN, Kwasny MJ, Kaiser SM, Gray EL, et al. A randomized noninferiority trial evaluating remotely-delivered stepped care for depression using internet cognitive behavioral therapy (CBT) and telephone CBT. Behav Res Ther 2019 Dec;123:103485 [FREE Full text] [doi: 10.1016/j.brat.2019.103485] [Medline: 31634738]
68. Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, et al. An internet-based intervention for depression in primary care in spain: a randomized controlled trial. J Med Internet Res 2016;18(8):e231 [FREE Full text] [doi: 10.2196/jmir.5695] [Medline: 27565118]
69. Nakao S, Nakagawa A, Oguchi Y, Mitsuda D, Kato N, Nakagawa Y, et al. Web-based cognitive behavioral therapy blended with face-to-face sessions for major depression: randomized controlled trial. J Med Internet Res 2018 Sep 21;20(9):e10743 [FREE Full text] [doi: 10.2196/10743] [Medline: 30249583]
70. Nyström MBT, Stenling A, Sjöström E, Neely G, Lindner P, Hassmén P, et al. Behavioral activation versus physical activity via the internet: a randomized controlled trial. J Affect Disord 2017 Dec;215:85-93 [FREE Full text] [doi: 10.1016/j.jad.2017.03.018] [Medline: 28319696]
71. Oehler C, Görges F, Rogalla M, Rummel-Kluge C, Hegerl U. Efficacy of a guided web-based self-management intervention for depression or dysthymia: randomized controlled trial with a 12-month follow-up using an active control condition. J Med Internet Res 2020 Jul 14;22(7):e15361 [FREE Full text] [doi: 10.2196/15361] [Medline: 32673233]
72. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol Med 2014 Jun;44(8):1675-1689 [FREE Full text] [doi: 10.1017/S0033291713002092] [Medline: 24148703]
73. Perini S, Titov N, Andrews G. Clinician-assisted Internet-based treatment is effective for depression: randomized controlled trial. Aust N Z J Psychiatry 2009 Jun;43(6):571-578. [doi: 10.1080/00048670902873722] [Medline: 19440890]
74. Pfeiffer PN, Pope B, Houck M, Benn-Burton W, Zivin K, Ganoczy D, et al. Effectiveness of peer-supported computer-based CBT for depression among veterans in primary care. Psychiatr Serv 2020 Mar 01;71(3):256-262. [doi: 10.1176/appi.ps.201900283] [Medline: 31931686]
75. Pihlaja S, Lahti J, Lipsanen JO, Ritola V, Gummerus E, Stenberg J, et al. Scheduled telephone support for internet cognitive behavioral therapy for depression in patients at risk for dropout: pragmatic randomized controlled trial. J Med Internet Res 2020 Jul 23;22(7):e15732 [FREE Full text] [doi: 10.2196/15732] [Medline: 32706658]
76. Reins JA, Boß L, Lehr D, Berking M, Ebert DD. The more I got, the less I need? Efficacy of Internet-based guided self-help compared to online psychoeducation for major depressive disorder. J Affect Disord 2019 Mar 01;246:695-705. [doi: 10.1016/j.jad.2018.12.065] [Medline: 30611913]
77. Ren Z, Li X, Zhao L, Yu X, Li Z, Lai L, et al. Effectiveness and mechanism of internet-based self-help intervention for depression: the Chinese version of MoodGYM. Acta Psychologica Sinica 2016;48(7):818. [doi: 10.3724/sp.j.1041.2016.00818]

78. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. NPJ Digit Med 2020;3:85 [FREE Full text] [doi: 10.1038/s41746-020-0293-8] [Medline: 32566763]
79. Richards DA, Hill JJ, Gask L, Lovell K, Chew-Graham C, Bower P, et al. Clinical effectiveness of collaborative care for depression in UK primary care (CADET): cluster randomised controlled trial. BMJ 2013 Aug 19;347:f4913 [FREE Full text] [doi: 10.1136/bmj.f4913] [Medline: 23959152]
80. Rollman BL, Herbeck BB, Abebe KZ, Spring MB, Rotondi AJ, Rothenberger SD, et al. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: a randomized clinical trial. JAMA Psychiatry 2018 Jan 01;75(1):56-64. [doi: 10.1001/jamapsychiatry.2017.3379] [Medline: 29117275]
81. Rosso IM, Killgore WDS, Olson EA, Webb CA, Fukunaga R, Auerbach RP, et al. Internet-based cognitive behavior therapy for major depressive disorder: a randomized controlled trial. Depress Anxiety 2017 Mar;34(3):236-245 [FREE Full text] [doi: 10.1002/da.22590] [Medline: 28009467]
82. Sandoval LR, Buckey JC, Ainslie R, Tombari M, Stone W, Hegel MT. Randomized controlled trial of a computerized interactive media-based problem solving treatment for depression. Behav Ther 2017 May;48(3):413-425 [FREE Full text] [doi: 10.1016/j.beth.2016.04.001] [Medline: 28390503]
83. Schuver KJ, Lewis BA. Mindfulness-based yoga intervention for women with depression. Complement Ther Med 2016 Jun;26:85-91. [doi: 10.1016/j.ctim.2016.03.003] [Medline: 27261987]
84. Selmi PM, Klein MH, Greist JH, Sorrell SP, Erdman HP. Computer-administered cognitive-behavioral therapy for depression. Am J Psychiatry 1990 Jan;147(1):51-56. [doi: 10.1176/ajp.147.1.51] [Medline: 2403473]
85. Smith J, Newby JM, Burston N, Murphy MJ, Michael S, Mackenzie A, et al. Help from home for depression: a randomised controlled trial comparing internet-delivered cognitive behaviour therapy with bibliotherapy for depression. Internet Interv 2017 Sep;9:25-37 [FREE Full text] [doi: 10.1016/j.invent.2017.05.001] [Medline: 30135834]
86. Steinmann M, Heddaeus D, Liebherz S, Daubmann A, Härter M, Watzke B. Effectiveness of telephone-administered cognitive-behavioral psychotherapy for depression with versus without additional letters: a randomized controlled trial. Telemed J E Health 2020 Mar;26(3):347-353. [doi: 10.1089/tmj.2018.0311] [Medline: 31013466]
87. Thase ME, Wright JH, Eells TD, Barrett MS, Wisniewski SR, Balasubramani GK, et al. Improving the efficiency of psychotherapy for depression: computer-assisted versus standard CBT. Am J Psychiatry 2018 Mar 01;175(3):242-250. [doi: 10.1176/appi.ajp.2017.17010089] [Medline: 28969439]
88. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
89. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
90. Torkan H, Blackwell SE, Holmes EA, Kalantari M, Neshat-Doost HT, Maroufi M, et al. Positive imagery cognitive bias modification in treatment-seeking patients with major depression in Iran: a pilot study. Cognit Ther Res 2014;38:132-145 [FREE Full text] [doi: 10.1007/s10608-014-9598-8] [Medline: 24634554]
91. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: a randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
92. Watkins ER, Taylor RS, Byng R, Baeyens C, Read R, Pearson K, et al. Guided self-help concreteness training as an intervention for major depression in primary care: a Phase II randomized controlled trial. Psychol Med 2012 Jul;42(7):1359-1371 [FREE Full text] [doi: 10.1017/S0033291711002480] [Medline: 22085757]
93. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
94. Welch ES, Weigand A, Hooker JE, Philip NS, Tyrka AR, Press DZ, et al. Feasibility of computerized cognitive-behavioral therapy combined with bifrontal transcranial direct current stimulation for treatment of major depression. Neuromodulation 2019 Dec;22(8):898-903. [doi: 10.1111/ner.12807] [Medline: 30153360]
95. Williams AD, Blackwell SE, Mackenzie A, Holmes EA, Andrews G. Combining imagination and reason in the treatment of depression: a randomized controlled trial of internet-based cognitive-bias modification and internet-CBT for depression. J Consult Clin Psychol 2013 Oct;81(5):793-799 [FREE Full text] [doi: 10.1037/a0033247] [Medline: 23750459]
96. Williams AD, O'Moore K, Blackwell SE, Smith J, Holmes EA, Andrews G. Positive imagery cognitive bias modification (CBM) and internet-based cognitive behavioral therapy (iCBT): a randomized controlled trial. J Affect Disord 2015 Jun 01;178:131-141 [FREE Full text] [doi: 10.1016/j.jad.2015.02.026] [Medline: 25805405]
97. Wright JH, Wright AS, Albano AM, Basco MR, Goldsmith LJ, Raffield T, et al. Computer-assisted cognitive therapy for depression: maintaining efficacy while reducing therapist time. Am J Psychiatry 2005 Jun;162(6):1158-1164. [doi: 10.1176/appi.ajp.162.6.1158] [Medline: 15930065]

98. Zagorscak P, Heinrich M, Sommer D, Wagner B, Knaevelsrud C. Benefits of individualized feedback in internet-based interventions for depression: a randomized controlled trial. Psychother Psychosom 2018;87(1):32-45. [doi: 10.1159/000481515] [Medline: 29306945]
99. Zwerenz R, Becker J, Knickenberg RJ, Siepmann M, Hagen K, Beutel ME. Online self-help as an add-on to inpatient psychotherapy: efficacy of a new blended treatment approach. Psychother Psychosom 2017;86(6):341-350. [doi: 10.1159/000481177] [Medline: 29131090]
100. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
101. Massoudi B, Holvast F, Bockting CLH, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: a systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:728-743. [doi: 10.1016/j.jad.2018.11.050] [Medline: 30447572]
102. Ahern E, Kinsella S, Semkovska M. Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis. Expert Rev Pharmacoecon Outcomes Res 2018 Feb;18(1):25-41. [doi: 10.1080/14737167.2018.1407245] [Medline: 29145746]
103. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can J Psychiatry 2013 Jul;58(7):376-385. [doi: 10.1177/070674371305800702] [Medline: 23870719]
104. Cuijpers P, Turner EH, Mohr DC, Hofmann SG, Andersson G, Berking M, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. Psychol Med 2014 Mar;44(4):685-695. [doi: 10.1017/S0033291713000457] [Medline: 23552610]
105. Wentzel J, van der Vaart R, Bohlmeijer ET, 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;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
106. Baumeister H, Grässle C, Ebert DD, Krämer LV. Blended psychotherapy—verzahnte psychotherapie: das beste aus zwei welten? Psychotherapie Dialog 2018 Nov 28;19(04):33-38. [doi: 10.1055/a-0592-0264]
107. Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. Arch Intern Med 2006 Nov 27;166(21):2314-2321. [doi: 10.1001/archinte.166.21.2314] [Medline: 17130383]
108. Thota AB, Sipe TA, Byard GJ, Zometa CS, Hahn RA, McKnight-Eily LR, Community Preventive Services Task Force. Collaborative care to improve the management of depressive disorders: a community guide systematic review and meta-analysis. Am J Prev Med 2012 May;42(5):525-538. [doi: 10.1016/j.amepre.2012.01.019] [Medline: 22516495]
109. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
110. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 2016 Jul 12;6(7):e010247 [FREE Full text] [doi: 10.1136/bmjopen-2015-010247] [Medline: 27406637]
111. Beintner I, Vollert B, Zarski A, Bolinski F, Musiat P, Görlich D, et al. Adherence reporting in randomized controlled trials examining manualized multisession online interventions: systematic review of practices and proposal for reporting standards. J Med Internet Res 2019 Aug 15;21(8):e14181 [FREE Full text] [doi: 10.2196/14181] [Medline: 31414664]
112. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of Internet interventions. Internet Interv 2014 Mar;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]

Abbreviations

cCBT: computerized cognitive behavioral therapy
f2f: face-to-face
ITT: intention-to-treat
PI: prediction interval
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
RR: risk ratio
SMD: standardized mean difference
TAU: treatment as usual
TBI: technology-based psychological intervention
TIDECA: Comparative Effectiveness of Technology-Based Interventions in Different Steps of Depression Care

Edited by G Eysenbach; submitted 25.09.20; peer-reviewed by ECY Su, A Gulliver, E Forsell; comments to author 01.12.20; revised version received 08.02.21; accepted 04.04.21; published 13.06.21.

Please cite as:
Köhnen M, Kriston L, Härter M, Baumeister H, Liebherz S
Effectiveness and Acceptance of Technology-Based Psychological Interventions for the Acute Treatment of Unipolar Depression: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(6):e24584
URL: https://www.jmir.org/2021/6/e24584/
doi:10.2196/24584
PMID:36260395

Review

Effectiveness of Virtual Reality Interventions for Adolescent Patients in Hospital Settings: Systematic Review

Brad Ridout[1], PhD, MAPS; Joshua Kelson[2], PhD; Andrew Campbell[1], PhD, MAPS; Kate Steinbeck[3], PhD

[1]Cyberpsychology Research Group, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[2]Faculty of Business, Justice, and Behavioural Sciences, Charles Sturt University, Bathurst, Australia

[3]Discipline of Child and Adolescent Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

Corresponding Author:
Brad Ridout, PhD, MAPS
Cyberpsychology Research Group
Faculty of Medicine and Health
The University of Sydney
City Road
Sydney, 2006
Australia
Phone: 61 2 8627 1444
Email: brad.ridout@sydney.edu.au

Abstract

Background: Given the high level of interest and increasing familiarity with virtual reality among adolescents, there is great potential to use virtual reality to address adolescents' unique health care delivery needs while in hospital. While there have been reviews on the use of virtual reality for specific health conditions and procedures, none to date have reviewed the full scope of virtual reality hospital interventions for adolescents who are often combined with children as a homogenous group, despite the fact that adolescents experience virtual environments different from children.

Objective: The aim of this review was to systematically identify available evidence regarding the use of virtual reality interventions for adolescent patients in hospital settings to evaluate effectiveness, suitability, and safety and identify opportunities for future research.

Methods: PubMed, PsycINFO, Medline, and Scopus databases were searched using keywords and phrases. Retrieved abstracts (n=1525) were double screened, yielding 276 articles for full-text screening. Of these, 8 articles met inclusion criteria. Data were extracted to a standardized coding sheet, and a narrative synthesis was performed due to the heterogeneity of the studies.

Results: Four RCTs and 4 single-case reports were identified for inclusion, all of which aimed to reduce pain or anxiety. The scenarios targeted were burn pain, venipuncture, chemotherapy, preoperative anxiety, and palliative care. Three out of 4 RCTs found significant reductions in pain or anxiety outcomes measures when using virtual reality compared to standard care or other distraction techniques; however, only 1 study combined self-reported experiences of pain or anxiety with any physiological measures. Single-case reports relied primarily upon qualitative feedback, with patients reporting reduced pain or anxiety and a preference for virtual reality to no virtual reality.

Conclusions: Virtual reality can provide a safe and engaging way to reduce pain and anxiety in adolescents while in hospital, particularly when virtual reality software is highly immersive and specifically designed for therapeutic purposes. As VR becomes more accessible and affordable for use in hospitals, larger and more diverse studies that capitalize on adolescents' interest in and aptitude for virtual reality, and on the full range of capabilities of this emerging technology, are needed to build on these promising results.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020198760; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020198760

(J Med Internet Res 2021;23(6):e24967) doi:10.2196/24967

KEYWORDS

virtual reality; hospital; pain; anxiety; adolescents

Introduction

Interest in the use of virtual reality (VR) in the health sector has increased steadily over the past decade; recent advances have made VR technology more immersive, flexible, portable, and affordable. VR has been studied in the treatment of a wide range of clinical conditions, including pain management [1-3], rehabilitation [4], anxiety [5], phobias [6], and posttraumatic stress disorder [7].

The term *VR* has been used in health research to refer to a variety of simulated experiences between an individual and a 3D computer-generated environment, including videogames using a standard monitor [8]. However, VR is now mostly considered to require immersion [9,10] in a 3D environment that provides the user with the illusory experience of being in a place other than where they physically are (known as presence) [11,12]. This is usually achieved with a stereoscopic head-mounted display, often with motion tracking that allows the user to actively determine their field of view (by moving their head) to interact with the environment.

The ability of VR to modulate subjective experience lends itself to use by patients in hospital settings, where it may be used to offer respite from stressful or confining environments, such as hospital wards or emergency departments, or as a distraction from chronic or procedural pain or anxiety [13]. While there have been reviews into particular applications of VR for specific health conditions or in-hospital procedures (eg, burn management [14], procedural pain [1]), to date there has only been 1 systematic review, which included only RCTs, on the full scope of immersive VR use in hospital settings [13]. Furthermore, reviews [1,13-15] have typically combined results for patient groups ranging from young children to older adults and have not taken into consideration differences in levels of enthusiasm, aptitude, nor predisposition toward VR between these populations [16]. Brain-imaging research has suggested that young children's brains process virtual environments different from the manner in which adolescent and adult brains process virtual environments [17-19].

VR is particularly appealing to adolescents; a recent US survey found that 73% of adolescents aged 11 to 15 years are fairly to extremely interested in VR [20]. Adolescents, defined by the World Health Organization as aged between 10 and 19 years [21], are developmentally distinct from young children and adults in terms of neurocognitive and physical maturation [22], yet studies of the use of VR for pediatric patients to date have primarily used samples that combined adolescents with young children [23-26], which makes it difficult to determine adolescent-specific outcomes.

Hospitals may be particularly stressful environments for adolescents, who are at a vulnerable stage in their development [27]. Their health issues require different responses from the hospital system than those young children [28,29] or adults [30] require, because they are affected by the physical, emotional, psychological, and sociocultural stages of adolescence as they develop their identity and autonomy [31]. For example, adolescents with chronic health conditions have preferences and concepts of care that differ from those of adults [32,33]. Adolescents also have more awareness and knowledge of their health than younger children yet may not have the emotional or cognitive resources to deal with their situation as well as older populations [34]. Furthermore, adolescents present hospitals with unique medical and psychological challenges, such as those related to the onset of mental health disorders [35], and difficulties in ensuring compliance [36], especially when adolescents perceive their dignity to be violated [37]. It is for these reasons that a growing body of research has advocated for specialist physician training and accreditation in adolescent medicine [38,39] and for adolescents to be studied as a developmentally distinct group separate from children and adults [30].

Given the high level of interest and familiarity with VR among adolescents and VR's increasing affordability and accessibility, there is great potential to use VR to address the unique health care delivery needs of adolescents while in hospital, both as inpatients and outpatients. It is, therefore, important to understand how VR has been used in the treatment of adolescents in hospital settings to date and whether their interest in this technology translates to its enthusiastic use for therapeutic purposes, so that researchers and physicians can leverage the potential health benefits of VR for this population.

To date, no systematic review has been conducted on the overall use of VR in hospital settings among adolescents. This systematic review is therefore needed to determine how VR is currently being used to improve the well-being and experiences of adolescents in hospital; evaluate the effectiveness, suitability, and safety of such interventions; and identify opportunities for future research.

Methods

Search Strategy

This systematic review was performed using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses [40]) (Multimedia Appendix 1). PubMed, PsycINFO, Medline, and Scopus databases were searched using the following search phrase: (virtual reality OR VR) AND (adolescen* OR child* OR pediatric* OR youth OR teen*) AND (hospital* OR inpatient OR treat* OR surg*). Searches were conducted in May 2020 and restricted to English-language articles published in peer-reviewed journals between January 2005 and May 2020. This review (and protocol) was registered (PROSPERO; CRD42020198760).

The database searches yielded 2214 records (549 from PubMed, 216 from PsycINFO, 592 from Medline, and 857 from Scopus), from which 689 duplicates were removed. Manual searches of previous reviews, key journals, and reference lists of key articles were conducted; however, no additional records were identified.

Eligibility Criteria

Record Type

As VR research in health is still in its infancy and recent reviews of restricted populations have identified mostly exploratory and feasibility studies [41,42], empirical studies of all research designs (including single-case reports) were included, to ensure the review was comprehensive.

Participants

Studies with a target population of adolescents, defined as aged between 10 to 19 years [21], were included.

Intervention Hardware

This review was restricted to studies that used immersive forms of VR delivered through a head-mounted display. Because it is not always possible for patients using VR in health settings to move their head, this review included VR studies that were immersive but not necessarily interactive (ie, passive forms of VR such as prerecorded immersive VR videos were included). Although Cave Automatic Virtual Environment systems, which use projections to display the VR environment on walls of a special-purpose room, are considered immersive, they are far less affordable and widely used than head-mounted displays [10,43] and are seldom available in hospital settings. Studies using Cave Automatic Virtual Environment systems were, therefore, excluded, along with studies that used nonimmersive hardware such as computer monitors.

Intervention Settings

Studies on interventions that took place in hospital settings were included. Participants could be either inpatients or outpatients.

Screening Process

The screening process involved 2 stages: (1) title and abstract exclusion and (2) full-text exclusion. All records were independently screened by 2 reviewers (BR and JK) in both stages to establish relevance for inclusion. Any discrepancies between reviewers' decisions were resolved by discussion with the authors who were not involved in selection (AC and KS) until mutual agreement was reached.

Data Extraction

Data from studies included in the review were extracted by 1 reviewer (BR) to a standardized coding sheet, which was then checked by a second reviewer (JK). Data extracted for synthesis were reference source (first author surname; year of publication), methodology (health problem or procedure under investigation; study design; risk of bias assessment), participants (country; attrition rate; sample size, age, and gender characteristics), intervention details (treatment conditions; VR equipment and software), and findings (primary outcomes, ie, impact on physical or psychological measures; usability outcomes, ie, measures of engagement, acceptability).

Risk of Bias Assessment

Risk of bias was assessed independently by 2 reviewers (BR and JK) for RCTs, using the Cochrane Collaboration tool [44], and for single-case reports, using the Methodological Quality and Synthesis of Case Series and Case Reports Protocol [45]. Any discrepancies between reviewers were resolved by discussion. Given that it is not possible to blind participants or personnel to a VR intervention condition, the questions about performance bias and detection bias were not assessed for RCTs. An additional domain—whether any confounder variables at baseline were accounted for—was included under other risks. For case reports, causality questions ("were other alternative causes that may explain the observation ruled out?"; "was there a challenge/rechallenge phenomenon?"; "was there a dose–response effect?") were not applicable to this study and were not assessed.

Data Analysis

Given the small number of studies included in this review and the heterogeneity of their aims, research design, and outcome measures, quantitative meta-analysis was not possible. A narrative synthesis approach was used to describe the findings of this systematic review.

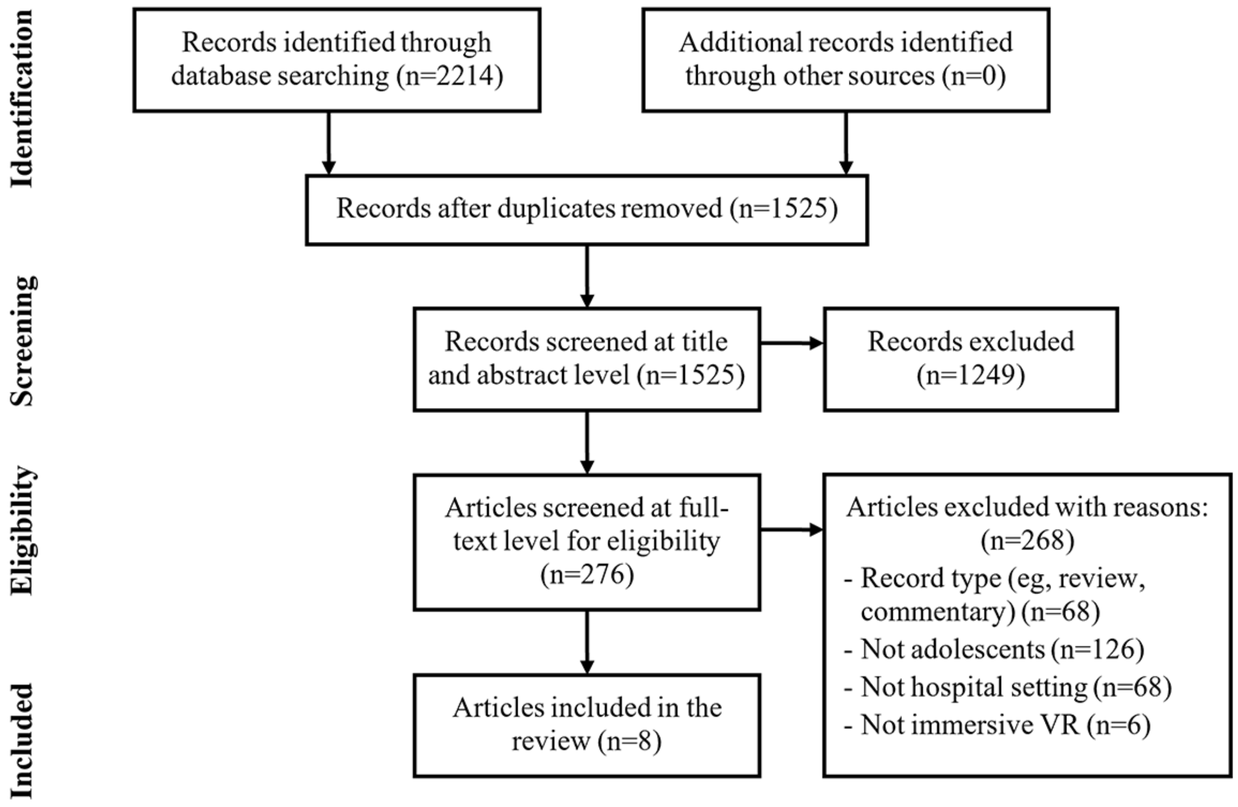
Results

Study Selection

The literature search yielded 2214 records. Of the 1525 records that remained after duplicate removal, 1249 were excluded because their titles and abstracts indicated that they were not about virtual reality interventions for adolescent patients in hospital settings, leaving 276 articles to be assessed for eligibility based on inclusion criteria.

During the second stage of screening, 68 articles were excluded based on record type; 126 were excluded because the target population was not limited to adolescents; 68 articles were excluded because the intervention did not take place in hospital settings; and 6 were excluded because they did not use immersive VR. A total of 8 studies were included (Figure 1).

Figure 1. Selection flowchart.



Health Problems/Procedures and Study Designs

Detailed characteristics of the included articles (n=8) are provided in Multimedia Appendix 2. Four studies [46-49] were RCTs and 4 studies [50-53] were single-case reports. Of the RCTs, 2 studies [46,49] compared VR to standard care, 1 study [47] compared VR to standard distraction (TV, music, books), and 1 study [48] compared VR to 2 control groups—standard care and passive distraction (watching a movie). The study design of 1 RCT [46] was incorrectly described as quasi-experimental and the control group was described as waitlist; however, private communication with the authors confirmed that while on the waitlist, the control group received the same standard care as the intervention group but without the VR component (G Manshaee, PhD, email communication, August 9, 2020). Given this information and the fact that group assignment was randomized, we classified the study as an RCT.

Three studies used VR for distraction from burn pain (2 RCTs [47,48], 1 case report [50]), 2 studies used VR for distraction from pain and anxiety during chemotherapy sessions (1 RCT [46], 1 case report [51]), 1 study [49] used VR for distraction from pain and anxiety during venipuncture, 1 case report used VR to reduce preoperative anxiety [52], and 1 case report used VR to improve well-being in palliative care [53].

Three of the RCTs [46,48,49] compared pretest and posttest measures of outcome variables (1 study also included 7-day and 1-month follow-ups [46]), while 1 RCT [47] measured burn pain intensity, observed pain behaviors, and objective physiological indicators (heart rate and oxygen saturation) at 3 time points: baseline, during dressing removal, and during dressing application. Two case reports [50,51] compared outcome variables measured with and without VR, while the other 2 case reports [52,53] only recorded qualitative assessments of the patients' VR experiences.

Risk of Bias Assessment

The risk of bias for the RCT studies (Multimedia Appendix 3) was low for most studies in most domains. Baseline confounding variables were adequately accounted for in 2 RCTs [47,49], but there was a high risk of bias for this domain in 1 RCT [46] and an unclear risk in the other RCT [48]. All of the case reports were assessed as having some risk of methodological bias (Multimedia Appendix 4). None of the case reports described clear selection methods, and 2 case reports [52,53] did not adequately ascertain outcomes (ie, results were qualitative or broadly descriptive only).

Participant Characteristics

Studies were conducted in the United States (n=5), Australia (n=1), Iran (n=1), and Canada (n=1). The RCTs investigating VR for burn pain had similar sample sizes (28 participants [48] and 41 participants [47]) and gender distributions (approximately one-third of participants were female). In the RCT investigating VR and chemotherapy [46], all the participants were female (sample size of 30). The RCT investigating VR and venipuncture had a large sample size (143 participants) and an even gender distribution [49]. The overall attrition rate in RCTs was negligible, with 2 studies [46,47] reporting no attrition. In the other 2 studies, attrition occurred primarily due to participants withdrawing or rescheduling prior to their procedures.

The age range for all included studies fell within the World Health Organization definition of 10 to 19 years [21], with the exception of 1 RCT [49], which also included 16 participants (16/143, 11.2%) aged 20 or 21 years who were defined as adolescents in the study because they were being treated at a Children's Hospital. The age distributions of 2 RCTs were similar (range 11-17 years, mean 13.1 years [47]; range 10-17 years, mean 13.5 years [48]), while 1 RCT [46] had an age range of 14 to 18 years (mean 14.9 years). In the case reports, 3 patients were aged between 10 and 12 years (2 males [50,52] and 1 female [53]), and 1 patient was a 17-year-old male who was morbidly obese [51].

VR Software

Most studies tested VR programs that were specifically designed for therapeutic purposes: SnowWorld (University of Washington Harborview Burn Center and Firsthand Technology Inc), the first VR software created specifically for pain distraction during burn wound redressing [50]; Bear Blast (AppliedVR Inc), a fast-paced interactive game designed for pain distraction, for venipuncture [49]; and Ocean Journey, a passive (ie, noninteractive) VR therapy film, for chemotherapy [46]. In 1 case report [53], the patient, who was in palliative care, was provided with a range of 360° videos (Wishplay), designed to allow patients to have experiences beyond the limitations of their illness or disability, which included a figure-skating experience custom-made for the patient.

Three studies used off-the-shelf software not designed for therapeutic purposes: 1 case report on a patient undergoing chemotherapy used Ocean Descent (Sony Interactive Entertainment); 1 RCT on burn pain used age-appropriate interactive VR games (Chicken Little for 11- to 13-year-olds; Need for Speed for 14- to 17-year-olds) [47]; and 1 case report used Oculus Rooms to connect the patient suffering preoperative anxiety to their parent (who was located in the preoperative area) and allowed them to play a virtual board game together while the patient was transported to the operating table and until they lost consciousness from the anesthesia [52]. This case report was the only study included that used software that allowed virtual interaction with other users.

The VR session length in half of the studies [46,50-52] was between 20 and 30 minutes, in line with recommendations for VR session length [54]. The exceptions were the study with venipuncture [49], in which the session length was less than 5 minutes due to the short procedure; the palliative care study (5-10 minutes per video) [53], and the 2 studies with burn wound care [47,48], in which session length varied greatly depending on the length of time required to remove and apply dressings (2-100 minutes). A facilitator was present in all sessions to help guide the adolescent participants, with the exception of the palliative care study [53], in which the patient was first guided to ensure tolerability, and then left to use the VR device at their own leisure for approximately 4 weeks.

VR Equipment

Studies used a variety of immersive VR devices (Multimedia Appendix 2). The studies with patients with burn wounds, which were the oldest studies included, each used different VR headsets connected to desktop computers with user control via joystick or trackball (ie, no head tracking): Kipping et al [47] excluded patients with burn wound locations that would impact their ability to wear a head-mounted display, while Jeffs et al [48] and Hoffman et al [50] used head-mounted displays on custom-built adjustable arm devices to allow patients with burn wounds on the head to be included. The other studies used widely available consumer head-mounted displays with head tracking capabilities (Oculus Go, Oculus Rift, Sony PlayStation VR), including 3 studies [46,49,53] that utilized smartphones as the display and processor (Samsung VR Gear, Merge VR Goggles, Google Daydream).

Primary Measures and Outcomes

There was almost no commonality in the measures used to assess primary outcomes (Multimedia Appendix 2), with the exception of the use of a visual analog scale to assess pain in RCTs for venipuncture [49] and burn pain [47].

Gold and Mahrer [49] found that using VR for distraction during venipuncture resulted in significantly less procedural pain and procedural anxiety (each measured with visual analog scales), and significantly better affect (measured with the Facial Affect scale) for VR than that for standard care, when controlling for baseline pain. Secondary analyses revealed that patients with high anxiety sensitivity benefited the most from VR [49].

Kipping et al [47] measured burn pain intensity (with a visual analog scale) and observed pain behaviors, heart rate, and oxygen saturation, but the only significant difference between VR (using off-the-shelf interactive games) and standard distraction was fewer pain behaviors observed by nurses during dressing removal for the VR condition. Jeffs et al [48] found that procedural pain while using SnowWorld was significantly lower for VR (and was the only condition in which procedural pain was reduced compared to preprocedure pain) than that for passive distraction or standard care (adjusted for age, sex, preprocedure pain, state anxiety, opiate use, and treatment length). A patient with burns undergoing occupational therapy also reported lower pain when using SnowWorld than that experienced in standard occupational therapy sessions the days before and after the VR session [50].

Sharifpour et al [46] found that, after 8 weekly 30-minute chemotherapy sessions while watching the noninteractive VR film Ocean Journey, patients reported lower pain intensity, pain anxiety, pain catastrophizing, and pain self-efficacy compared to the standard care control group. These effects were maintained for subsequent weekly chemotherapy sessions without VR, at 7-day and 1-month follow-ups. One case report with chemotherapy [51] did not measure pain or anxiety directly; the use of VR during a monthly lumbar puncture procedure (injection of intrathecal chemotherapy) reduced the amount of analgesics and anxiolytics required by the patient by approximately half compared to the amount required in the previous 4 monthly procedures without VR, and procedure and recovery were significantly faster (42% and 30%, respectively).

Two case reports [51,52] did not use quantitative measures. In 1 study [52], staff reported that the patient remained calm and showed no signs of distress or anxiety while using VR during transportation to the operating therapy and while being administered anesthesia. In the other study [51], the patient in palliative care reported that using VR distracted them from their pain and loneliness and had a positive impact on their mood.

Usability Measures and Outcomes

Three of the RCTs [47-49] used quantitative usability or engagement measures, but there was no commonality among them. Kipping et al [47] found self-reported presence (with a visual analog scale) while using VR to be positive but with room for improvement (mean 6.1 out of 10). Jeffs et al [48] found that engagement, using a nonstandard question about perceived ability to pay attention to the distraction (either VR or passive distraction) rated on a 5-point scale, demonstrated a significant negative correlation with both anxiety and procedural pain reduction in both groups, though a direct comparison between VR and passive distraction groups was not performed. The venipuncture RCT [49] measured usability using 2 investigator-developed Likert-type scale measures and a qualitative questionnaire: results indicated a high level of immersion and satisfaction with the VR game, and 92% of participants reported no feelings of sickness during the VR session. Jeffs et al [48] was the only other RCT to mention side-effects, with none reported. The chemotherapy RCT [46] did not report any usability or engagement data.

All case reports found the VR experiences to be both immersive and fun, with no feelings of sickness or discomfort reported; however, only 1 study (the patient in palliative care [53]) confirmed an absence of side-effects. The only usability issues that were reported were from the patient suffering preoperative anxiety (orientation disturbance) within Oculus Rooms: while the patient was being moved during transportation to the operating room (although the patient was able to quickly correct the orientation without assistance); and poor Wi-Fi connectivity [52].

Discussion

Principal Findings

The aim of this systematic review was to identify studies that investigated the use of immersive VR using head-mounted displays to improve the well-being and experiences of adolescents in hospital settings. We chose this age group and setting for the following reasons: First, it is often assumed that adolescents will be enthusiastic users of therapeutic digital technology because they are engaged users of similar technology for social and leisure purposes; however, there is little research evidence to date to support that this is a reasonable assumption. Second, adolescents are often viewed as challenging or hyperemotional in binary (pediatric or adult) health care settings [35,36,55]. Part of this hyperemotionality is related to neurocognitive development [56] and may be better managed with distraction or immersion using VR than by traditional models of care used for younger and older populations.

We identified 8 eligible articles (4 RCTs and 4 case reports), all of which aimed to reduce pain or anxiety. The number of articles was low, but not unremarkably so, given that research with adolescents in hospital settings is often stalled around consent and risk issues [57], and in studies, adolescents are often combined with young children or adults [57]. The health problems and procedures targeted were burn pain, venipuncture, chemotherapy, preoperative anxiety, and palliative care. While the lack of large RCTs precluded any meta-analysis, most [46,48,49] found significant reductions in pain or anxiety outcomes measures when using VR compared to those when using standard care or other distraction techniques. There was little commonality in the measures used to evaluate primary and usability outcomes, with only 1 study [47] combining self-reported pain or anxiety with physiological measures. Risk of bias was generally low for the RCTs but was relatively high for the single-case reports due to their study design and unclear selection method. Caution needs to be exercised when interpreting these findings, particularly for the case reports.

VR was well received by adolescents, who generally found it to be safe, fun, immersive, and engaging. The attrition rates in the RCT studies were very low compared to those of other VR studies [54], which supports the view that adolescents find VR more engaging than other populations, which is highly relevant.

A wide range of head-mounted displays were used, including headsets that utilized smartphones as the display and processor. Studies [48,50] also demonstrated the efficacy of using custom-mounted head-mounted displays combined with hand controllers for patients who are undergoing treatments that prevent them from moving their head. Consistent with the findings of previous studies with pediatric patients [58-61], there were very few reports of side-effects associated with VR use; however, only 1 study [49] in this review quantitatively measured feelings of nausea.

Therapeutic Mechanisms of VR Use by Adolescents

The synthesis of findings from this review present an up-to-date account of the evidence base for VR use by adolescents in hospitals. There are sound reasons why VR might be generally an effective therapeutic intervention for pain, anxiety, and other distress associated with necessary health care situations, as well as some reasons why VR in these health care situations might be particularly advantageous for adolescents.

The mechanism by which VR reduces the experience of pain and anxiety has mostly been attributed to active distraction. That is, it directs the patient's attention away from their treatment or condition by requiring them to interact with the VR environment. The neuromatrix theory of pain [62] suggests that the perception and experience of pain can be altered by cognitive, sensory, and affective experiences. VR interventions that actively distract by engaging cognitive resources (by being engaging and interactive), offer high sensory stimulation (by being immersive), and lead to positive affect (by being enjoyable) can therefore reduce the neurological resources available for processing pain [63,64]. Studies have shown that adolescents are more sensitive to pain and likely to become more emotionally dysregulated when faced with situations that are unfamiliar [65], so active distraction using VR may be particularly beneficial for adolescents, especially in hospital settings.

It has previously been demonstrated that active distraction using VR is particularly effective in reducing burn pain [14,66], especially when using SnowWorld, a game specifically designed for burns patients. Consistent with these findings, 2 studies [48,50] in this review that used SnowWorld reported a reduction in adolescents' experiences of pain. In contrast, a study [47] that used off-the-shelf VR games for burn pain found that they were not significantly better than using standard distraction methods and induced only a moderate level of presence. Previous studies [67-69] have shown that the level of presence or immersion experienced by patients in VR interventions is directly correlated with the level of pain reduction. Kipping and colleagues [47] noted that any savings realized from using off-the-shelf VR software are likely to be at the expense of effectiveness.

While passive or less immersive scenarios have been shown to provide little relief from severe pain, for example, in patients with burns [70], these scenarios have been effective in chemotherapy. Sharifpour and colleagues [46] demonstrated that watching a passive VR film during weekly 30-minute chemotherapy sessions was not only effective in reducing adolescents' scores on a range of pain and anxiety measures, but also that this effect was maintained for subsequent weekly sessions without VR. These findings suggest a mechanism other than active distraction in this study. It has been shown that VR can lead to increased cognitive control over pain, by facilitating relaxation and changing the way people think about pain [71,72], specifically by reducing pain catastrophizing, and increasing pain self-efficacy (ie, the ability to tolerate and control pain). This would fit with the concept that situations facilitating mastery and self-control are positive learning experiences for adolescents [73]. This suggests that the use of VR in the management of pain during chemotherapy, chronic pain associated with cancer, or other medical conditions may provide lasting benefits, even afterward when one is not using a VR device [74].

The patient in the palliative care case report was also experiencing chronic pain, but reported another therapeutic mechanism—immersive VR helped as a distraction from the loneliness and boredom associated with long-term hospitalization and waiting for health care attention [53]. With emerging research showing that VR use in palliative care settings is acceptable and well tolerated [75,76], more studies are needed to investigate expanding the use of VR to settings beyond distraction and relieving boredom, such as connecting adolescents with their peers and family to share a VR experience or play VR games together. Isolation from friends and peers is one of the most frequently reported negative aspects of hospitalization for adolescents [27,30]. Using VR to play games with friends or communicate with family via home-to-hospital live streaming [77] would therefore be particularly suitable for improving the mood and well-being of hospitalized adolescents, especially given their high level of enthusiasm and predisposition toward VR [16,20] and VR's emerging acceptability for use in these settings [26].

Implications for Future Research

The studies included in this systematic review suggest that there is tremendous potential for immersive VR to improve the well-being of adolescents in hospital settings. While the evidence base for this specific population is not yet established, the inclusion of case studies in this review demonstrates feasibility for several new applications of VR for adolescents in hospital settings and provides researchers with directions for potential interventions for the future.

This review also highlighted a number of methodological concerns that researchers in this field should seek to address in future studies. These include inconsistency across studies in the selection of primary outcome and usability measures, a lack physiological measures to complement subjective measures, and the difficulty in replicating studies due to the wide use of customized software and hardware.

Another contribution of this adolescent-specific review is the identification of several gaps in the literature for this population. The studies included in this review focused on pain and pain-related anxiety, however, recent studies in nonhospital settings have shown great potential for VR to be used to reduce other kinds of psychological distress in adolescents in a range of circumstances [78-82]. Emergency departments are a hospital environment that adolescents are more likely to use than other age groups [83], but also one that adolescents find particularly distressing due to long waits, loud sounds, bright lights, privacy intrusions, and exposure to the distress of others [84]. VR could therefore be effective in not only dealing with acute procedural pain while in emergency [15], but also in blocking out distressing stimuli for a calming experience that could assist them in regulating their emotions [79]. A recent study has demonstrated that VR use in the emergency department can significantly reduce levels of anxiety and anger in adults [85]; therefore, future studies should test whether these findings are also true in adolescents.

There is also opportunity for greater use of VR to connect adolescents in hospital with others in real time. Gupta and colleagues [52] demonstrated the feasibility of using Oculus Rooms to connect with a parent to relieve preoperative anxiety, but there are many other situations in which hospitalized adolescents are separated from their family and peers (eg, isolation rooms, palliative care, because of visiting hours limitations or having travelled from rural and remote areas). While the use of VR headsets that use smartphones as the display and processor (eg, Samsung VR Gear) is declining due to the increase in development and affordability of all-inclusive head-mounted displays (eg, Oculus Quest 2) [86-88], the use of either type of head-mounted display may be well-suited for hospitalized adolescents, who could potentially use their own smartphones to connect with others in VR, if given access to reliable Wi-Fi and the right software apps.

Future studies should seek to expand on the findings of the case reports included in this review, by empirically investigating the feasibility of using VR to connect adolescents with others to not only distract from pain and anxiety, but also improve well-being by relieving distress and boredom when hospitalized or waiting in emergency departments for extended periods. The potential for VR to improve pain self-efficacy and to better cope with chronic pain should also be explored.

Limitations

This systematic review was limited to studies on the use of immersive VR to improve the well-being and experiences of adolescents in hospital settings. As such, studies that combined adolescents with other populations such as younger children or adults were excluded. Searches were also limited to English language publications and excluded grey literature and conference papers, as we were concerned with identifying the state of peer-reviewed research. Given the small number of studies identified, and the heterogeneity of aims, research designs, and outcome measures, it was not possible to conduct a meta-analysis.

While 8 studies may seem to be a low number for a systematic review, this is not uncommon when reviewing novel uses of emerging technology in specific settings and populations [89-92]. Given that VR devices have only recently become more affordable and portable (indicated by half of the included studies being published from 2018 onward), and hence more suitable for wide implementation in hospital settings, it is likely that more research in this area is currently in progress and has yet to be published. Alternatively, the challenges of undertaking novel research in busy hospital settings cannot be discounted as a reason for the low number of studies that were found.

The strengths of this review include a clear research question, prospectively registered protocol, thorough search strategy with more than 1 assessor, and the inclusion of all research designs to capture applications of VR at various stages of development. No prior reviews have specifically investigated the full range of immersive VR use to improve the in-hospital experiences of adolescents.

Conclusion

This was the first systematic review of published studies on the use of immersive VR to improve the well-being and experiences of adolescents in hospital settings without confounds from younger children and adults. This is an important contribution to the field of VR health research, given that adolescents are developmentally distinct from other age groups, and present hospitals with unique challenges and health care delivery needs for which VR may be a useful and appealing tool. Studies varied in terms of quality and design, from RCTs to single-case reports that support the feasibility of potential interventions for the future. Overall, there was support for the effectiveness of VR in hospitals to reduce pain and anxiety in adolescents, particularly when VR software was highly immersive and specifically designed for therapeutic purposes.

There were examples of both active and passive distraction mechanisms being effectively used by VR interventions, although the latter is currently underutilized for adolescents in hospital settings. There was also a lack of RCTs investigating the effect of VR on adolescents in hospital without combining results with younger children, who process virtual environments differently. Future studies should use larger sample sizes and RCT designs, evaluate physiological and psychological outcome measures in addition to self-report measures, and address current gaps in the literature by empirically investigating the use of VR to relieve psychological distress in adolescents while in hospital, connect adolescents with friends and family to improve their well-being, and help adolescents develop skills to better tolerate and control chronic pain.

The use of VR in the health sector has enormous potential, especially for use with adolescents, who have a keen interest and aptitude toward this emerging treatment modality. As VR technology continues to improve and become more affordable, the evidence base for its effectiveness in relieving adolescent pain and distress in hospital settings should continue to grow.

Acknowledgments

This project was supported by the Wellbeing Health & Youth Centre of Research Excellence in Adolescent Health, funded by the National Health and Medical Research Council.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 29 KB - jmir_v23i6e24967_app1.docx]

Multimedia Appendix 2
Study characteristics.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v23i6e24967_app2.xlsx]

Multimedia Appendix 3
Risk of bias - randomized controlled trials.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i6e24967_app3.xlsx]

Multimedia Appendix 4
Risk of bias - case reports.
[XLSX File (Microsoft Excel File), 10 KB - jmir_v23i6e24967_app4.xlsx]

References

1. Chan E, Foster S, Sambell R, Leong P. Clinical efficacy of virtual reality for acute procedural pain management: a systematic review and meta-analysis. PLoS One 2018 Jul 27;13(7):e0200987 [FREE Full text] [doi: 10.1371/journal.pone.0200987] [Medline: 30052655]
2. Eijlers R, Utens EMWJ, Staals LM, de Nijs PFA, Berghmans JM, Wijnen RMH, et al. Systematic review and meta-analysis of virtual reality in pediatrics: effects on pain and anxiety. Anesth Analg 2019 Nov;129(5):1344-1353 [FREE Full text] [doi: 10.1213/ANE.0000000000004165] [Medline: 31136330]
3. Malloy KM, Milling LS. The effectiveness of virtual reality distraction for pain reduction: a systematic review. Clin Psychol Rev 2010 Dec;30(8):1011-1018. [doi: 10.1016/j.cpr.2010.07.001] [Medline: 20691523]
4. Howard MC. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Comput Hum Behav 2017 May;70:317-327. [doi: 10.1016/j.chb.2017.01.013]
5. Carl E, Stein AT, Levihn-Coon A, Pogue JR, Rothbaum B, Emmelkamp P, et al. Virtual reality exposure therapy for anxiety and related disorders: a meta-analysis of randomized controlled trials. J Anxiety Disord 2019 Jan;61:27-36. [doi: 10.1016/j.janxdis.2018.08.003] [Medline: 30287083]
6. Botella C, Fernández-Álvarez J, Guillén V, García-Palacios A, Baños R. Recent progress in virtual reality exposure therapy for phobias: a systematic review. Curr Psychiatry Rep 2017 Jul 24;19(7):42. [doi: 10.1007/s11920-017-0788-4] [Medline: 28540594]
7. Deng W, Hu D, Xu S, Liu X, Zhao J, Chen Q, et al. The efficacy of virtual reality exposure therapy for PTSD symptoms: a systematic review and meta-analysis. J Affect Disord 2019 Oct 01;257:698-709. [doi: 10.1016/j.jad.2019.07.086] [Medline: 31382122]
8. Levac DE, Miller PA. Integrating virtual reality video games into practice: clinicians' experiences. Physiother Theory Pract 2013 Oct 30;29(7):504-512. [doi: 10.3109/09593985.2012.762078] [Medline: 23362843]
9. Wiederhold B, Bouchard S. Advances in Virtual Reality and Anxiety Disorders. Boston, MA: Springer US; 2014.
10. Howard MC. Virtual reality interventions for personal development: a meta-analysis of hardware and software. Hum Comput Interact 2018 May 15;34(3):205-239. [doi: 10.1080/07370024.2018.1469408]
11. Slater M, Usoh M. Presence in immersive virtual environments. 1993 Presented at: IEEE Virtual Reality Annual International Symposium; September 18-22; Washington, DC. [doi: 10.1109/vrais.1993.380793]
12. Skarbez R, Brooks, Jr. FP, Whitton MC. A survey of presence and related concepts. ACM Comput Surv 2018 Jan 12;50(6):1-39. [doi: 10.1145/3134301]
13. Dascal J, Reid M, IsHak W, Spiegel B, Recacho J, Rosen B, et al. Virtual reality and medical inpatients: a systematic review of randomized, controlled trials. Innov Clin Neurosci 2017;14(1-2):14-21 [FREE Full text] [Medline: 28386517]
14. Scapin S, Echevarría-Guanilo ME, Boeira Fuculo Junior PR, Gonçalves N, Rocha PK, Coimbra R. Virtual reality in the treatment of burn patients: a systematic review. Burns 2018 Sep;44(6):1403-1416. [doi: 10.1016/j.burns.2017.11.002] [Medline: 29395400]
15. Pancekauskaitė G, Jankauskaitė L. Paediatric pain medicine: pain differences, recognition and coping acute procedural pain in paediatric emergency room. Medicina (Kaunas) 2018 Nov 27;54(6):94 [FREE Full text] [doi: 10.3390/medicina54060094] [Medline: 30486427]
16. Burch A. VR and consumer sentiment. Touchstone VR. 2016. URL: https://touchstoneresearch.com/vr-and-consumer-sentiment/ [accessed 2020-01-14]
17. Baumgartner T, Valko L, Esslen M, Jäncke L. Neural correlate of spatial presence in an arousing and noninteractive virtual reality: an EEG and psychophysiology study. Cyberpsychol Behav 2006 Feb;9(1):30-45. [doi: 10.1089/cpb.2006.9.30] [Medline: 16497116]
18. Baumgartner T, Speck D, Wettstein D, Masnari O, Beeli G, Jäncke L. Feeling present in arousing virtual reality worlds: prefrontal brain regions differentially orchestrate presence experience in adults and children. Front Hum Neurosci 2008;2:8 [FREE Full text] [doi: 10.3389/neuro.09.008.2008] [Medline: 18958209]
19. Deutsch J, McCoy S. Virtual reality and serious games in neurorehabilitation of children and adults: prevention, plasticity and participation. 2016 Presented at: Section on Pediatrics of the American Physical Therapy Association; November 9-10; Denver, CO p. S23-S36. [doi: 10.1097/pep.0000000000000387]
20. Yamada-Rice D, Mushtaq F, Woodgate A, Bosmans D, Douthwaite A, Douthwaite I, et al. Children and virtual reality: emerging possibilities and challenges. DigLitEY. 2017. URL: http://digilitey.eu/wp-content/uploads/2015/09/CVR-Final-PDF-reduced-size.pdf [accessed 2020-01-14]
21. Adolescent health. World Health Organization. 2020. URL: https://www.who.int/health-topics/adolescent-health/ [accessed 2020-01-14]
22. Meeus W. Adolescent Development: Longitudinal Research into the Self, Personal Relationships and Psychopathology. New York, NY: Routledge; 2019.

23. Piskorz J, Czub M. Effectiveness of a virtual reality intervention to minimize pediatric stress and pain intensity during venipuncture. J Spec Pediatr Nurs 2018 Jan 20;23(1):e12201. [doi: 10.1111/jspn.12201] [Medline: 29155488]
24. Schmitt YS, Hoffman HG, Blough DK, Patterson DR, Jensen MP, Soltani M, et al. A randomized, controlled trial of immersive virtual reality analgesia, during physical therapy for pediatric burns. Burns 2011 Feb;37(1):61-68 [FREE Full text] [doi: 10.1016/j.burns.2010.07.007] [Medline: 20692769]
25. Hua Y, Qiu R, Yao W, Zhang Q, Chen X. The effect of virtual reality distraction on pain relief during dressing changes in children with chronic wounds on lower limbs. Pain Manag Nurs 2015 Oct;16(5):685-691. [doi: 10.1016/j.pmn.2015.03.001] [Medline: 25972074]
26. Flujas-Contreras J, Ruiz-Castañeda D, Gómez I. Promoting emotional well-being in hospitalized children and adolescents with virtual reality: usability and acceptability of a randomized controlled trial. Comput Inform Nurs 2020 Feb;38(2):99-107. [doi: 10.1097/CIN.0000000000000586] [Medline: 31895058]
27. Massimo L, Rossoni N, Mattei F, Bonassi S, Caprino D. Needs and expectations of adolescent in-patients: the experience of Gaslini Children's Hospital. Int J Adolesc Med Health 2016 Feb;28(1):11-17. [doi: 10.1515/ijamh-2014-0070] [Medline: 25720045]
28. Kullgren KA, Sullivan SK, Bravender T. Understanding the unique needs of hospitalized adolescents and young adults referred for psychology consults. Clin Pediatr (Phila) 2018 Oct 07;57(11):1286-1293. [doi: 10.1177/0009922818774339] [Medline: 29732921]
29. Sawyer S, Ambresin A, Bennett K, Hearpes S, Romaniuk P. Towards and adolescent friendly hospital. The Royal Children's Hospital. Sydney, Australia: Centre for Adolescent Health, The Royal Children's Hospital; 2011. URL: https://www.rch.org.au/uploadedFiles/Main/Content/cah/2012%20Adolscent%20Report.pdf [accessed 2020-04-01]
30. Sturrock T, Masterson L, Steinbeck K. Adolescent appropriate care in an adult hospital: the use of a youth care plan. Aust J Adv Nurs 2007;24(3):49-53. [Medline: 17518166]
31. Jamalimoghadam N, Yektatalab S, Momennasab M, Ebadi A, Zare N. How do hospitalized adolescents feel safe? a qualitative study. J Nurs Res 2019 Apr 01;27(2):e14. [doi: 10.1097/jnr.0000000000000285] [Medline: 30102645]
32. Farrant B, Watson P. Health care delivery: perspectives of young people with chronic illness and their parents. J Paediatr Child Health 2004 Apr;40(4):175-179. [doi: 10.1111/j.1440-1754.2004.00333.x] [Medline: 15009544]
33. Jedeloo S, van Staa A, Latour JM, van Exel NJA. Preferences for health care and self-management among Dutch adolescents with chronic conditions: a Q-methodological investigation. Int J Nurs Stud 2010 May;47(5):593-603. [doi: 10.1016/j.ijnurstu.2009.10.006] [Medline: 19900675]
34. Shulman EP, Smith AR, Silva K, Icenogle G, Duell N, Chein J, et al. The dual systems model: Review, reappraisal, and reaffirmation. Dev Cogn Neurosci 2016 Feb;17:103-117 [FREE Full text] [doi: 10.1016/j.dcn.2015.12.010] [Medline: 26774291]
35. Moses T. Determinants of mental illness stigma for adolescents discharged from psychiatric hospitalization. Soc Sci Med 2014 May;109:26-34. [doi: 10.1016/j.socscimed.2013.12.032] [Medline: 24695363]
36. Michaud P, Schrier L, Ross-Russel R, van der Heijden L, Dossche L, Copley S, et al. Paediatric departments need to improve residents' training in adolescent medicine and health: a position paper of the European Academy of Paediatrics. Eur J Pediatr 2018 Apr 21;177(4):479-487. [doi: 10.1007/s00431-017-3061-2] [Medline: 29270826]
37. Jamalimoghadam N, Yektatalab S, Momennasab M, Ebadi A, Zare N. Hospitalized adolescents' perception of dignity: a qualitative study. Nurs Ethics 2019 May;26(3):728-737. [doi: 10.1177/0969733017720828] [Medline: 28805113]
38. Kanbur N, Akgül S, Merrick J. Adolescent health and medicine: a global perspective on training adolescent health professionals. Int J Adolesc Med Health 2016 Aug 01;28(3):229-230. [doi: 10.1515/ijamh-2016-5001] [Medline: 26115490]
39. Sawyer S, Farrant B, Hall A, Kennedy A, Payne D, Steinbeck K, et al. Adolescent and young adult medicine in Australia and New Zealand: towards specialist accreditation. Int J Adolesc Med Health 2016 Aug 01;28(3):253-261. [doi: 10.1515/ijamh-2016-5006] [Medline: 26115493]
40. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
41. Naylor M, Ridout B, Campbell A. A scoping review identifying the need for quality research on the use of virtual reality in workplace settings for stress management. Cyberpsychol Behav Soc Netw 2020 Aug 01;23(8):506-518. [doi: 10.1089/cyber.2019.0287] [Medline: 32486836]
42. Tuena C, Pedroli E, Trimarchi PD, Gallucci A, Chiappini M, Goulene K, et al. Usability issues of clinical and research applications of virtual reality in older people: a systematic review. Front Hum Neurosci 2020 Apr 8;14:93. [doi: 10.3389/fnhum.2020.00093]
43. Slater M, Sanchez-Vives MV. Enhancing our lives with immersive virtual reality. Front Robot AI 2016 Dec 19;3:74. [doi: 10.3389/frobt.2016.00074]
44. Higgins J, Altman D. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011).: The Cochrane Collaboration; 2011. URL: handbook.cochrane.org [accessed 2020-06-30]
45. Murad MH, Sultan S, Haffar S, Bazerbachi F. Methodological quality and synthesis of case series and case reports. BMJ Evid Based Med 2018 Apr 02;23(2):60-63 [FREE Full text] [doi: 10.1136/bmjebm-2017-110853] [Medline: 29420178]

46. Sharifpour S, Manshaee G, Sajjadian I. Effects of virtual reality therapy on perceived pain intensity, anxiety, catastrophising and self-efficacy among adolescents with cancer. Couns Psychother Res 2020 Mar 28;21(1):218-226 [FREE Full text] [doi: 10.1002/capr.12311]
47. Kipping B, Rodger S, Miller K, Kimble RM. Virtual reality for acute pain reduction in adolescents undergoing burn wound care: a prospective randomized controlled trial. Burns 2012 Aug;38(5):650-657. [doi: 10.1016/j.burns.2011.11.010] [Medline: 22348801]
48. Jeffs D, Dorman D, Brown S, Files A, Graves T, Kirk E, et al. Effect of virtual reality on adolescent pain during burn wound care. J Burn Care Res 2014;35(5):395-408. [doi: 10.1097/BCR.0000000000000019] [Medline: 24823326]
49. Gold J, Mahrer N. Is virtual reality ready for prime time in the medical space? a randomized control trial of pediatric virtual reality for acute procedural pain management. J Pediatr Psychol 2018 Apr 01;43(3):266-275. [doi: 10.1093/jpepsy/jsx129] [Medline: 29053848]
50. Hoffman HG, Meyer WJ, Ramirez M, Roberts L, Seibel EJ, Atzori B, et al. Feasibility of articulated arm mounted Oculus Rift Virtual Reality goggles for adjunctive pain control during occupational therapy in pediatric burn patients. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):397-401 [FREE Full text] [doi: 10.1089/cyber.2014.0058] [Medline: 24892204]
51. Thomas JJ, Albietz J, Polaner D. Virtual reality for lumbar puncture in a morbidly obese patient with leukemia. Paediatr Anaesth 2018 Nov 04;28(11):1059-1060. [doi: 10.1111/pan.13505] [Medline: 30284748]
52. Gupta A, Joseph Thomas J. Use of virtual reality as a surrogate for parental presence during anesthetic induction: a case report. A A Pract 2019 Dec 15;13(12):454-456. [doi: 10.1213/XAA.0000000000001114] [Medline: 31703005]
53. Weingarten K, Macapagal F, Parker D. Virtual reality: endless potential in pediatric palliative care: a case report. J Palliat Med 2020 Jan;23(1):147-149. [doi: 10.1089/jpm.2019.0207] [Medline: 31170022]
54. Benbow AA, Anderson PL. A meta-analytic examination of attrition in virtual reality exposure therapy for anxiety disorders. J Anxiety Disord 2019 Jan;61:18-26. [doi: 10.1016/j.janxdis.2018.06.006] [Medline: 30646997]
55. Findlay S, Pinzon J, Goldberg E, Frappier J. Issues of care for hospitalized youth. Paediatr Child Health 2008 Jan;13(1):61-68 [FREE Full text] [doi: 10.1093/pch/13.1.61] [Medline: 19119356]
56. Guyer AE, Silk JS, Nelson EE. The neurobiology of the emotional adolescent: from the inside out. Neurosci Biobehav Rev 2016 Nov;70:74-85 [FREE Full text] [doi: 10.1016/j.neubiorev.2016.07.037] [Medline: 27506384]
57. Steinbeck K, Baur L, Cowell C, Pietrobelli A. Clinical research in adolescents: challenges and opportunities using obesity as a model. Int J Obes (Lond) 2009 Jan;33(1):2-7. [doi: 10.1038/ijo.2008.263] [Medline: 19065147]
58. Morris L, Louw Q, Grimmer-Somers K. The effectiveness of virtual reality on reducing pain and anxiety in burn injury patients: a systematic review. Clin J Pain 2009;25(9):815-826. [doi: 10.1097/AJP.0b013e3181aaa909] [Medline: 19851164]
59. Ravi D, Kumar N, Singhi P. Effectiveness of virtual reality rehabilitation for children and adolescents with cerebral palsy: an updated evidence-based systematic review. Physiotherapy 2017 Sep;103(3):245-258. [doi: 10.1016/j.physio.2016.08.004] [Medline: 28109566]
60. Indovina P, Barone D, Gallo L, Chirico A, De Pietro G, Giordano A. Virtual reality as a distraction intervention to relieve pain and distress during medical procedures. Clin J Pain 2018 Sep;34(9):858-877. [doi: 10.1097/AJP.0000000000000599] [Medline: 29485536]
61. Faber AW, Patterson DR, Bremer M. Repeated use of immersive virtual reality therapy to control pain during wound dressing changes in pediatric and adult burn patients. J Burn Care Res 2013;34(5):563-568 [FREE Full text] [doi: 10.1097/BCR.0b013e3182777904] [Medline: 23970314]
62. Melzack R. Pain and the neuromatrix in the brain. J Dent Educ 2001 Dec;65(12):1378-1382. [Medline: 11780656]
63. Hoffman HG, Chambers GT, Meyer WJ, Arceneaux LL, Russell WJ, Seibel EJ, et al. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med 2011 Apr 25;41(2):183-191 [FREE Full text] [doi: 10.1007/s12160-010-9248-7] [Medline: 21264690]
64. Li A, Montaño Z, Chen VJ, Gold JI. Virtual reality and pain management: current trends and future directions. Pain Manag 2011 Mar;1(2):147-157 [FREE Full text] [doi: 10.2217/pmt.10.15] [Medline: 21779307]
65. Kozlowska K. Functional somatic symptoms in childhood and adolescence. Curr Opin Psychiatry 2013 Sep;26(5):485-492. [doi: 10.1097/YCO.0b013e3283642ca0] [Medline: 23867659]
66. Ahmadpour N, Keep M, Janssen A, Rouf AS, Marthick M. Design strategies for virtual reality interventions for managing pain and anxiety in children and adolescents: scoping review. JMIR Serious Games 2020 Jan 31;8(1):e14565 [FREE Full text] [doi: 10.2196/14565] [Medline: 32012042]
67. Hoffman HG, Seibel EJ, Richards TL, Furness TA, Patterson DR, Sharar SR. Virtual reality helmet display quality influences the magnitude of virtual reality analgesia. J Pain 2006 Nov;7(11):843-850. [doi: 10.1016/j.jpain.2006.04.006] [Medline: 17074626]
68. Gutierrez-Maldonado J, Gutierrez-Martinez O, Cabas-Hoyos K. Interactive and passive virtual reality distraction: effects on presence and pain intensity. Stud Health Technol Inform 2011;167:69-73. [Medline: 21685644]
69. Johnson S, Coxon M. Sound can enhance the analgesic effect of virtual reality. R Soc Open Sci 2016 Mar;3(3):150567 [FREE Full text] [doi: 10.1098/rsos.150567] [Medline: 27069646]

70. Phelan I, Furness P, Fehily O, Thompson A, Babiker N, Lamb M, et al. A mixed-methods investigation into the acceptability, usability, and perceived effectiveness of active and passive virtual reality scenarios in managing pain under experimental conditions. J Burn Care Res 2019 Jan 01;40(1):85-90. [doi: 10.1093/jbcr/iry052] [Medline: 30247616]
71. Shiri S, Feintuch U, Weiss N, Pustilnik A, Geffen T, Kay B, et al. A virtual reality system combined with biofeedback for treating pediatric chronic headache--a pilot study. Pain Med 2013 May 01;14(5):621-627. [doi: 10.1111/pme.12083] [Medline: 23659372]
72. Loreto-Quijada D, Gutiérrez-Maldonado J, Nieto R, Gutiérrez-Martínez O, Ferrer-García M, Saldaña C, et al. Differential effects of two virtual reality interventions: distraction versus pain control. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):353-358. [doi: 10.1089/cyber.2014.0057] [Medline: 24892197]
73. Casey B, Galván A, Somerville LH. Beyond simple models of adolescence to an integrated circuit-based account: a commentary. Dev Cogn Neurosci 2016 Feb;17:128-130 [FREE Full text] [doi: 10.1016/j.dcn.2015.12.006] [Medline: 26739434]
74. Gupta A, Scott K, Dukewich M. Innovative technology using virtual reality in the treatment of pain: does it reduce pain via distraction, or is there more to it? Pain Med 2018 Jan 01;19(1):151-159. [doi: 10.1093/pm/pnx109] [Medline: 29025113]
75. Johnson T, Bauler L, Vos D, Hifko A, Garg P, Ahmed M, et al. Virtual reality use for symptom management in palliative care: a pilot study to assess user perceptions. J Palliat Med 2020 Sep 01;23(9):1233-1238. [doi: 10.1089/jpm.2019.0411] [Medline: 31895637]
76. Niki K, Okamoto Y, Maeda I, Mori I, Ishii R, Matsuda Y, et al. Virtual reality use for symptom management in palliative care: a pilot study to assess user perceptions. J Palliat Med 2019 Jun;22(6):702-707. [doi: 10.1089/jpm.2018.0527] [Medline: 30676847]
77. Bakker A, Janssen L, Noordam C. Home to hospital live streaming with virtual reality goggles: a qualitative study exploring the experiences of hospitalized children. JMIR Pediatr Parent 2018 Dec 13;1(2):e10 [FREE Full text] [doi: 10.2196/pediatrics.9576] [Medline: 31518293]
78. Bossenbroek R, Wols A, Weerdmeester J, Lichtwarck-Aschoff A, Granic I, van Rooij MMJW. Efficacy of a virtual reality biofeedback game (DEEP) to reduce anxiety and disruptive classroom behavior: single-case study. JMIR Ment Health 2020 Mar 24;7(3):e16066 [FREE Full text] [doi: 10.2196/16066] [Medline: 32207697]
79. Hadley W, Houck C, Brown L, Spitalnick J, Ferrer M, Barker D. Moving beyond role-play: evaluating the use of virtual reality to teach emotion regulation for the prevention of adolescent risk behavior within a randomized pilot trial. J Pediatr Psychol 2019 May 01;44(4):425-435 [FREE Full text] [doi: 10.1093/jpepsy/jsy092] [Medline: 30551157]
80. Kahlon S, Lindner P, Nordgreen T. Virtual reality exposure therapy for adolescents with fear of public speaking: a non-randomized feasibility and pilot study. Child Adolesc Psychiatry Ment Health 2019;13:47 [FREE Full text] [doi: 10.1186/s13034-019-0307-y] [Medline: 31890004]
81. Kahlon S, Lindner P, Nordgreen T. Virtual reality exposure therapy for adolescents with fear of public speaking: a non-randomized feasibility and pilot study. Child Adolesc Psychiatry Ment Health 2019;13:47 [FREE Full text] [doi: 10.1186/s13034-019-0307-y] [Medline: 31890004]
82. Rowe LS, Jouriles EN, McDonald R. Reducing sexual victimization among adolescent girls: a randomized controlled pilot trial of my voice, my choice. Behav Ther 2015 May;46(3):315-327. [doi: 10.1016/j.beth.2014.11.003] [Medline: 25892168]
83. Jenkins B, Katz I. Adolescents' and young adults' use of Emergency Departments: an evidence check rapid review brokered by the Sax Institute for the NSW Kids and Families. Sax Institute. 2015 May. URL: https://www.saxinstitute.org.au/wp-content/uploads/Adolescents-and-young-adults-use-of-emergency-departments.pdf [accessed 2020-06-30]
84. Gerson R, Malas N, Feuer V, Silver G, Prasad R, Mroczkowski M. Erratum: This Article Corrects: "Best Practices for Evaluation and Treatment of Agitated Children and Adolescents (BETA) in the Emergency Department: Consensus Statement of the American Association for Emergency Psychiatry". West J Emerg Med 2019 Jul 2;20(4):688-689 [FREE Full text] [doi: 10.5811/westjem.2019.4.44160] [Medline: 31316711]
85. Sikka N, Shu L, Ritchie B, Amdur RL, Pourmand A. Virtual reality-assisted pain, anxiety, and anger management in the emergency department. Telemed J E Health 2019 Dec 01;25(12):1207-1215. [doi: 10.1089/tmj.2018.0273] [Medline: 30785860]
86. Robertson A. Phone-based VR is officially over. The Verge. 2019. URL: https://www.theverge.com/2019/10/16/20915791/google-daydream-samsung-oculus-gear-vr-mobile-vr-platforms-dead [accessed 2020-06-30]
87. Mosso Vázquez JL, Mosso Lara D, Mosso Lara JL, Miller I, Wiederhold MD, Wiederhold BK. Pain distraction during ambulatory surgery: virtual reality and mobile devices. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):15-21. [doi: 10.1089/cyber.2017.0714] [Medline: 30256662]
88. Cipresso P, Giglioli IAC, Raya MA, Riva G. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. Front Psychol 2018;9:2086 [FREE Full text] [doi: 10.3389/fpsyg.2018.02086] [Medline: 30459681]
89. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244 [FREE Full text] [doi: 10.2196/12244] [Medline: 30563811]
90. Rollin A, Ridout B, Campbell A. Digital health in melanoma posttreatment care in rural and remote Australia: systematic review. J Med Internet Res 2018 Sep 24;20(9):e11547 [FREE Full text] [doi: 10.2196/11547] [Medline: 30249578]

91. Larsen C, Oestergaard J, Ottesen B, Soerensen J. The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. Acta Obstet Gynecol Scand 2012 Sep;91(9):1015-1028 [FREE Full text] [doi: 10.1111/j.1600-0412.2012.01482.x] [Medline: 22693954]
92. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
VR: virtual reality

Edited by G Eysenbach; submitted 13.10.20; peer-reviewed by I Bell, V Rocío; comments to author 02.12.20; revised version received 09.12.20; accepted 15.04.21; published 28.06.21.

Please cite as:
Ridout B, Kelson J, Campbell A, Steinbeck K
Effectiveness of Virtual Reality Interventions for Adolescent Patients in Hospital Settings: Systematic Review
J Med Internet Res 2021;23(6):e24967
URL: https://www.jmir.org/2021/6/e24967
doi:10.2196/24967
PMID:34185015

Review

Health Recommender Systems: Systematic Review

Robin De Croon[1], PhD; Leen Van Houdt[1], MSc; Nyi Nyi Htun[1], PhD; Gregor Štiglic[2], PhD; Vero Vanden Abeele[1], PhD; Katrien Verbert[1], PhD

[1]Department of Computer Science, KU Leuven, Leuven, Belgium
[2]Faculty of Health Sciences, University of Maribor, Maribor, Slovenia

Corresponding Author:
Robin De Croon, PhD
Department of Computer Science
KU Leuven
Celestijnenlaan 200A
Leuven, 3001
Belgium
Phone: 32 16373976
Email: robin.decroon@kuleuven.be

Abstract

Background: Health recommender systems (HRSs) offer the potential to motivate and engage users to change their behavior by sharing better choices and actionable knowledge based on observed user behavior.

Objective: We aim to review HRSs targeting nonmedical professionals (laypersons) to better understand the current state of the art and identify both the main trends and the gaps with respect to current implementations.

Methods: We conducted a systematic literature review according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and synthesized the results. A total of 73 published studies that reported both an implementation and evaluation of an HRS targeted to laypersons were included and analyzed in this review.

Results: Recommended items were classified into four major categories: lifestyle, nutrition, general health care information, and specific health conditions. The majority of HRSs use hybrid recommendation algorithms. Evaluations of HRSs vary greatly; half of the studies only evaluated the algorithm with various metrics, whereas others performed full-scale randomized controlled trials or conducted in-the-wild studies to evaluate the impact of HRSs, thereby showing that the field is slowly maturing. On the basis of our review, we derived five reporting guidelines that can serve as a reference frame for future HRS studies. HRS studies should clarify who the target user is and to whom the recommendations apply, what is recommended and how the recommendations are presented to the user, where the data set can be found, what algorithms were used to calculate the recommendations, and what evaluation protocol was used.

Conclusions: There is significant opportunity for an HRS to inform and guide health actions. Through this review, we promote the discussion of ways to augment HRS research by recommending a reference frame with five design guidelines.

(J Med Internet Res 2021;23(6):e18035) doi:10.2196/18035

KEYWORDS

health recommender systems; recommender; recommendation system; health; health care; patient; layperson; systematic review; eHealth; evaluation; recommender technique; user interface; guidelines; mobile phone

Introduction

Research Goals

Current health challenges are often related to our modern way of living. High blood pressure, high glucose levels, and physical inactivity are all linked to a modern lifestyle characterized by sedentary living, chronic stress, or a high intake of energy-dense foods and recreational drugs [1]. Moreover, people usually make poor decisions related to their health for distinct reasons, for example, busy lifestyles, abundant options, and a lack of knowledge [2]. Practically, all modern lifestyle health risks are directly affected by people’s health decisions [3], such as an unhealthy diet or physical inactivity, which can contribute up to three-fourth of all health care costs in the United States [4]. Most risks can be minimized, prevented, or sometimes even reversed with small lifestyle changes. Eating healthily, increasing daily activities, and knowing where to find validated health information could lead to improved health status [5].

Health recommender systems (HRSs) offer the potential to motivate and engage users to change their behavior [6] and provide people with better choices and actionable knowledge based on observed behavior [7-9]. The overall objective of the HRS is to empower people to monitor and improve their health through technology-assisted, personalized recommendations. As one approach of modern health care is to involve patients in the cocreation of their own health, rather than just leaving it in the hands of medical experts [10], we limit the scope of this paper to HRSs that focus on laypersons, for example, nonhealth care professionals. These HRSs are different from clinical decision support systems that provide recommendations for health care professionals. However, laypersons also need to understand the rationale of recommendations, as echoed by many researchers and practitioners [11]. This paper also studies the role of a graphical user interface. To guide this study, we define our research questions (RQs) as follows:

RQ1: What are the main applications of the recent HRS, and what do these HRSs recommend?

RQ2: Which recommender techniques are being used across different HRSs?

RQ3: How are the HRSs evaluated, and are end users involved in their evaluation?

RQ4: Is a graphical user interface designed, and how is it used to communicate the recommended items to the user?

Recommender Systems and Techniques

Recommender techniques are traditionally divided into different categories [12,13] and are discussed in several state-of-the-art surveys [14]. *Collaborative filtering* is the most used and mature technique that compares the actions of multiple users to generate personalized suggestions. An example of this technique can typically be found on e-commerce sites, such as “Customers who bought this item also bought...” *Content-based filtering* is another technique that recommends items that are similar to other items preferred by the specific user. They rely on the characteristics of the objects themselves and are likely to be highly relevant to a user’s interests. This makes content-based filtering especially valuable for application domains with large libraries of a single type of content, such as MedlinePlus’ curated consumer health information [15]. *Knowledge-based filtering* is another technique that incorporates knowledge by logic inferences. This type of filtering uses explicit knowledge about an item, user preferences, and other recommendation criteria. However, knowledge acquisition can also be dynamic and relies on user feedback. For example, a camera recommender system might inquire users about their preferences, fixed or changeable lenses, and budget and then suggest a relevant camera. *Hybrid recommender systems* combine multiple filtering techniques to increase the accuracy of recommendation systems. For example, the *companies you may want to follow* feature in LinkedIn uses both content and collaborative filtering information [16]: collaborative filtering information is included to determine whether a company is similar to the ones a user already followed, whereas content information ensures whether the industry or location matches the interests of the user. Finally, recommender techniques are often augmented with additional methods to incorporate contextual information in the recommendation process [17], including recommendations via contextual prefiltering, contextual postfiltering, and contextual modeling [18].

HRSs for Laypersons

Ricci et al [12] define recommender systems as:

> *Recommender Systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user [13,19,20]. The suggestions relate to various decision-making processes, such as what items to buy, what music to listen to, or what online news to read.*

In this paper, we analyze how recommender systems have been used in health applications, with a focus on laypersons. Wiesner and Pfeifer [21] broadly define an HRS as:

> *a specialization of an RS [recommender system] as defined by Ricci et al [12]. In the context of an HRS, a recommendable item of interest is a piece of nonconfidential, scientifically proven or at least generally accepted medical information.*

Researchers have sought to consolidate the vast body of literature on HRSs by publishing several surveys, literature reviews, and state-of-the-art overviews. Table 1 provides an overview of existing summative studies on HRSs that identify existing research and shows the number of studies included, the method used to analyze the studies, the scope of the paper, and their contribution.

Table 1. An overview of the existing health recommender system overview papers.

Review	Papers, n	Method	Scope	Contribution
Sezgin and Özkan (2013) [22]	8	Systematic review	Provides an overview of the literature in 2013	Identifying challenges (eg, cyber-attacks, difficult integration, and data mining can cause ethical issues) and opportunities (eg, integration with personal health data, gathering user preferences, and increased consistency)
Calero Valdez et al (2016) [23]	17	Survey	Stresses the importance of the interface and HCI[a] of an HRS[b]	Providing a framework to incorporate domain understanding, evaluation, and specific methodology into the development process
Kamran and Javed (2015) [24]	7	Systematic review	Provides an overview of existing recommender systems with more focus on health care systems	Proposing a hybrid HRS
Afolabi et al (2015) [25]	22	Systematic review	Research empirical results and practical implementations of HRSs	Presenting a novel proposal for the integration of a recommender system into smart home care
Ferretto et al (2017) [26]	8	Systematic review	Identifies and analyzes HRSs available in mobile apps	Identifying HRSs that do not have many mobile health care apps
Hors-Fraile et al 2018 [27]	19	Systematic review	Identifies, categorizes, and analyzes existing knowledge on the use of HRSs for patient interventions	Proposing a multidisciplinary taxonomy, including integration with electronic health records and the incorporation of health promotion theoretical factors and behavior change theories
Schäfer et al (2017) [28]	24	Survey	Discusses HRSs to find personalized, complex medical interventions or support users with preventive health care measures	Identifying challenges subdivided into patient and user challenges, recommender challenges, and evaluation challenges
Sadasivam et al (2016) [29]	15	Systematic review	Research limitations of current CTHC[c] systems	Identifying challenges of incorporating recommender systems into CTHC. Proposing a future research agenda for CTHC systems
Wiesner and Pfeifer (2014) [21]	Not reported	Survey	Introduces HRSs and explains their usefulness to personal health record systems	Outlining an evaluation approach and discussing challenges and open issues
Cappella et al (2015) [30]	Not reported	Survey	Explores approaches to the development of a *recommendation system* for archives of public health messages	Reflecting on theory development and applications

[a] HCI: human-computer interaction.

[b] HRS: health recommender system.

[c] CTHC: computer-tailored health communication.

As can be seen in Table 1, the scope of the existing literature varies greatly. For example, Ferretto et al [26] focused solely on HRSs in mobile apps. A total of 3 review studies focused specifically on the *patient side* of the HRS: (1) Calero Valdez et al [23] analyzed the existing literature from a human-computer interaction perspective and stressed the importance of a good HRS graphical user interface; (2) Schäfer et al [28] focused on tailoring recommendations to end users based on health context, history, and goals; and (3) Hors-Fraile et al [27] focused on the individual user by analyzing how HRSs can target behavior change strategies. The most extensive study was conducted by Sadasivam et al [29]. In their study, most HRSs used knowledge-based recommender techniques, which might limit individual relevance and the ability to adapt in real time. However, they also reported that the HRS has the opportunity to use a near-infinite number of variables, which enables tailoring beyond designer-written rules based on data. The most important challenges reported were the cold start [31] where limited data are available at the start of the intervention, limited sample size, adherence, and potential unintended consequences [29]. Finally, we observed that these existing summative studies were often restrictive in their final set of papers.

Our contributions to the community are four-fold. First, we analyze a broader set of research studies to gain insights into the current state of the art. We do not limit the included studies to specific devices or patients in a clinical setting but *focus on laypersons* in general. Second, through a comprehensive analysis, we aim to identify the applications of recent HRS apps and gain insights into actionable knowledge that HRSs can provide to users (RQ1), to identify which recommender techniques have been used successfully in the domain (RQ2), how HRSs have been evaluated (RQ3), and the role of the user interface in communicating recommendations to users (RQ4). Third, based on our extensive literature review, we derive a reference frame with five reporting guidelines for future layperson HRS research. Finally, we collected and coded a unique data set of 73 papers, which is publicly available in Multimedia Appendix 1 [7-9,15,32-100] for other researchers.

Methods

Search Strategy

This study was conducted according to the key steps required for systematic reviews according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [101]. A literature search was conducted using the ACM Digital Library (n=2023), IEEExplore (n=277), and PubMed (n=93) databases. As mentioned earlier, in this systematic review we focused solely on HRSs with a focus on laypersons. However, many types of systems, algorithms, and devices can be considered as a HRS. For example, push notifications in a mobile health app or health tips prompted by web services can also be considered as health-related recommendations. To outline the scope, we limited the search terms to include a recommender or recommendation, as reported by the authors. The search keywords were as follows, using an inclusive OR: (*recommender* OR *recommendation systems* OR *recommendation system*) AND (*health* OR *healthcare* OR *patient* OR *patients*).

In addition, a backward search was performed by examining the bibliographies of the survey and review papers discussed in the Introduction section and the reference list of included studies to identify any additional studies. A forward search was performed to search for articles that cited the work summarized in Table 1.

Study Inclusion and Exclusion Criteria

As existing work did not include many studies (Table 1) and focused on a specific medical domain or device, such as mobile phones, this literature review used nonrestrictive inclusion criteria. Studies that met all the following criteria were included in the review: described an HRS whose primary focus was to improve health (eg, food recommenders solely based on user preferences [102] were not included); targeted laypersons (eg, activity recommendations targeted on a proxy user such as a coach [103] were not included); implemented the HRS (eg, papers describing an HRS concept are not included); reported an evaluation, either web-based or offline evaluation; peer-reviewed and published papers; published in English.

Papers were excluded when one of the following was true: the recommendations of HRSs were unclear; the full text was unavailable; or a newer version was already included.

Finally, when multiple papers described the same HRS, only the latest, relevant full paper was included.

Classification

To address our RQs, all included studies were coded for five distinct coding categories.

Study Details

To contextualize new insights, the publication year and publication venue were analyzed.

Recommended Items

HRSs are used across different health domains. To provide details on what is recommended, all papers were coded according to their respective health domains. To not limit the scope of potential items, no predefined coding table was used. Instead, all papers were initially coded by the first author. These resulting recommendations were then clustered together in collaboration with the coauthors into four categories, as shown in Multimedia Appendix 2.

Recommender Techniques

This category encodes the recommender techniques that were used: collaborative filtering [104], content-based filtering [105], knowledge-based filtering [106], and their hybridizations [107]. Some studies did not specify any algorithmic details or compared multiple techniques. Finally, when an HRS used contextual information, it was coded whether they used pre- or postfiltering or contextual modeling.

Evaluation Approach

This category encodes which evaluation protocols were used to measure the effect of HRSs. We coded whether the HRSs were evaluated through offline evaluations (no users involved), surveys, heuristic feedback from expert users, controlled user studies, deployments *in the wild*, and randomized controlled trials (RCTs). We also coded sample size and study duration and whether ethical approval was gathered and needed.

Interface and Transparency

Recommender systems are often perceived as a *black box*, as the rationale for recommendations is often not explained to end users. Recent research increasingly focuses on providing transparency to the inner logic of the system [11]. We encoded whether explanations are provided and, in this case, how such transparency is supported in the user interface. Furthermore, we also classified whether the user interface was designed for a specific platform, categorized as *mobile*, *web*, or other.

Data Extraction, Intercoder Reliability, and Quality Assessment

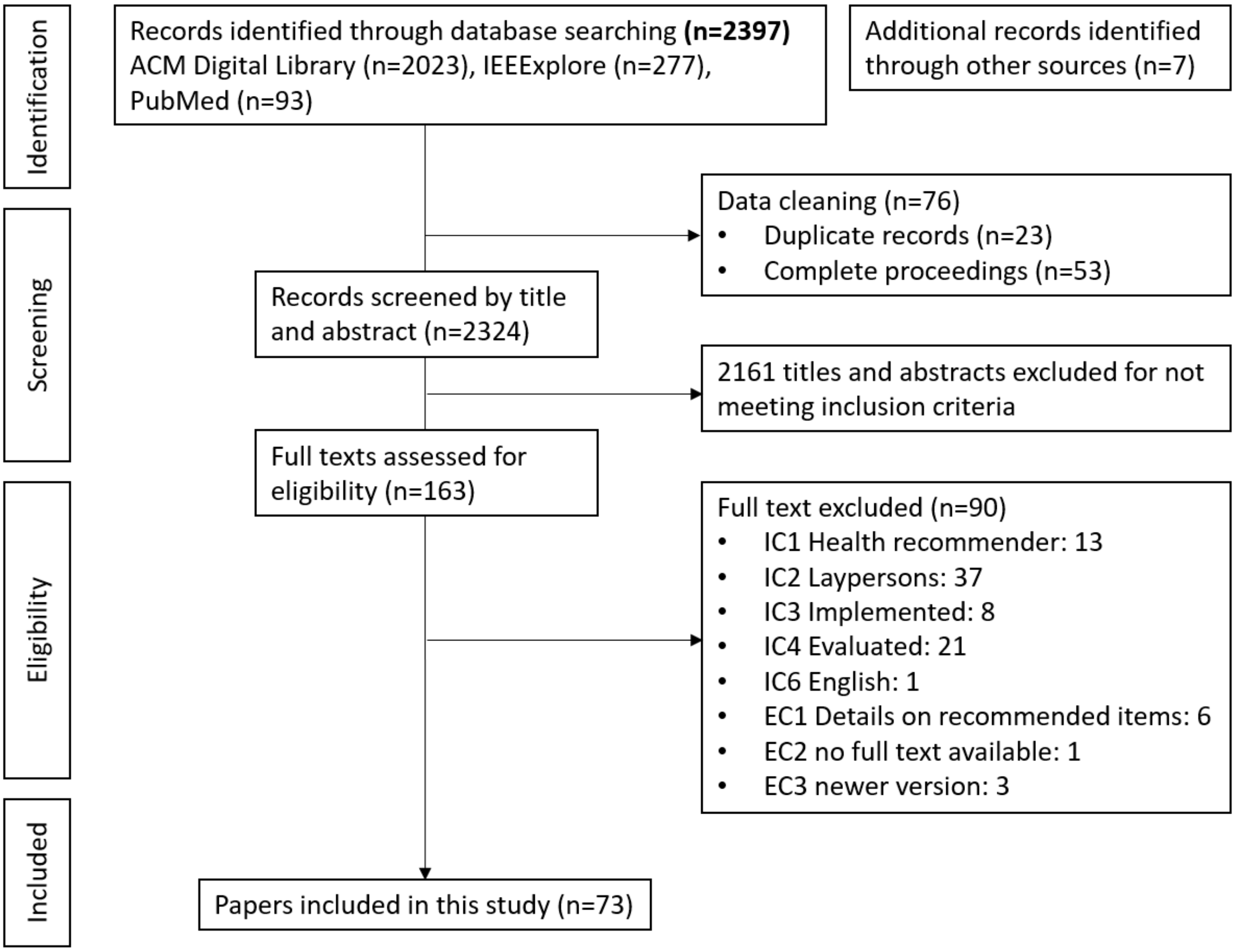
The required information for all included technologies and studies was coded by the first author using a data extraction form. Owing to the large variety of study designs, the included studies were assessed for quality (detailed scores given in Multimedia Appendix 1) using the tool by Hawker et al [108]. Using this tool, the *abstract and title*, *introduction and aims*, *method and data*, *sample size* (if applicable), *data analysis*, *ethics and bias*, *results*, *transferability or generalizability*, and *implications and usefulness* were allocated a score between 1 and 4, with higher scoring studies indicating higher quality. A random selection with 14% (10/73) of the papers was listed in a spreadsheet and coded by a second researcher following the defined coding categories and subcategories. The decisions made by the second researcher were compared with the first. With the recommended items (Multimedia Appendix 2), there was only one small disagreement between *physical activity* and *leisure activity* [32], but all other recommended items were rated exactly the same; the recommender techniques had a Cohen κ value of 0.71 (P<.001) and the evaluation approach scored a Cohen κ value of 0.81 (P<.001). There was moderate agreement (Cohen κ=0.568; P<.001) between the researchers concerning the quality of the papers. The interfaces used were in perfect agreement. Finally, the coding data are available in Multimedia Appendix 1.

Results

Study Details

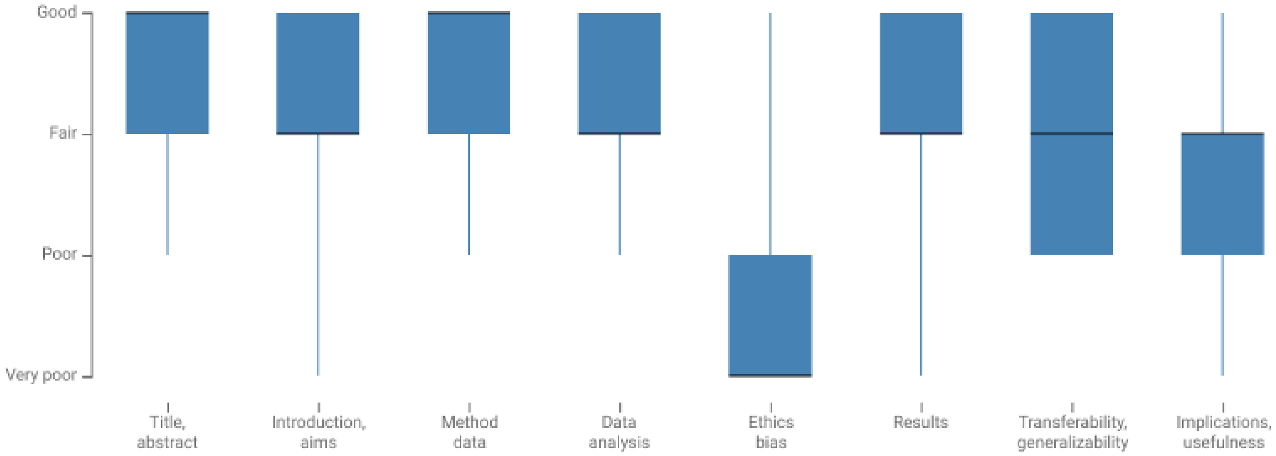
The literature in three databases yielded 2340 studies, of which only 23 were duplicates and 53 were full proceedings, leaving 2324 studies to be screened for eligibility. A total of 2161 studies were excluded upon title or abstract screening because they were unrelated to health or targeted at medical professionals or because the papers did not report an evaluation. Thus, the remaining 163 full-text studies were assessed for eligibility. After the removal of 90 studies that failed the inclusion criteria or met the exclusion criteria, 73 published studies remained. The search process is illustrated in Figure 1.

Figure 1. Flow diagram according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. EC: exclusion criteria; IC: inclusion criteria.



All included papers were published in 2009 or later, following an upward trend of increased popularity. The publication venues of HRSs are diverse. Only the PervasiveHealth [33-35], RecSys [36,37,109], and WI-IAT [38-40] conferences published 3 papers each that were included in this study. The *Journal of Medical Internet Research* was the only journal that occurred more frequently in our data set; 5 papers were published by *Journal of Medical Internet Research* [41-45]. The papers were first rated using Hawker tool [108]. Owing to a large number of offline evaluations, we did not include the sample score to enable a comparison between all included studies. The papers received an average score of 24.32 (SD 4.55, max 32; data set presented in Multimedia Appendix 1). Most studies scored *very poor* on reporting ethics and potential biases, as illustrated in Figure 2. However, there is an upward trend over the years in more adequate reporting of ethical issues and potential biases. The authors also limited themselves to their specific case studies and did not make any recommendations for policy (last box plot is presented in Figure 2). All 73 studies reported the use of different data sets. Although all recommended items were health related, only Asthana et al [46] explicitly mentioned using electronic health record data. Only 14% (10/73) [7,47-55] explicitly reported that they addressed the cold-start problem.

Figure 2. Distribution of the quality assessment using Hawker tool.



Recommended Items

Overview

Most HRSs operated in different domains and thus recommended different items. In this study, four nonmutually exclusive categories of recommended items were identified: lifestyle 33% (24/73), nutrition 36% (26/73), general health information 32% (23/73), and specific health condition–related recommendations 12% (9/73). The only significant trend we found is the increasing popularity of *nutrition advice*. Multimedia Appendix 2 shows the distribution of these recommended items.

Lifestyle

Many HRSs, 33% (24/73) of the included studies, suggest lifestyle-related items, but they differ greatly in their exact recommendations. Physical activity is often recommended. Physical activities are often personalized according to personal interests [56] or the context of the user [35]. In addition to physical activities, Kumar et al [32] recommend eating, shopping, and socializing activities. One study analyzes the data and measurements to be tracked for an individual and then recommends the appropriate wearable technologies to stimulate proactive health [46]. A total of 7 studies [7,9,42,53,57-59] more directly try to convince users to alter their behavior by recommending them to change, or alter their behavior: for example, Rabbi et al [7] learn “a user’s physical activity and dietary behavior and strategically suggests changes to those behaviors for a healthier lifestyle.” In another example, both Marlin et al [59] and Sadasivam et al [42] motivate users to stop smoking by providing them with tailored messages, such as “Keep in mind that cravings are temporary and will pass.” Messages could reflect the theoretical determinants of quitting, such as positive outcome expectations and self-efficacy enhancing small goals [42].

Nutrition

The influence of food on health is also clear from the large subset of HRSs dealing with nutrition recommendations. A mere 36% (26/73) of the studies recommend nutrition-related information, such as recipes [50], meal plans [36], restaurants [60], or even help with choosing healthy items from a restaurant menu [61]. Wayman and Madhvanath [37] provide automated, personalized, and goal-driven dietary guidance to users based on grocery receipt data. Trattner and Elsweiler [62] use postfiltering to focus on healthy recipes only and extended them with nutrition advice, whereas Ge et al [48] require users to first enter their preferences for better recommendations. Moreover, Gutiérrez et al [63] propose healthier alternatives through augmented reality when the users are shopping. A total of 7 studies specifically recommend healthy recipes [47,48,50,62,64-66]. Most HRSs consider the health condition of the user, such as the DIETOS system [67]. Other systems recommend recipes that are synthesized based on existing recipes and recommend new recipes [64], assist parents in making appropriate food for their toddlers [47], or help users to choose allergy-safe recipes [65].

General Health Information

According to 32% (23/73) of the included studies, providing access to trustworthy health care information is another common objective. A total of 5 studies focused on personalized, trustworthy information per se [15,55,68-70], whereas 5 others focused on guiding users through health care forums [52,71-74]. In total, 3 studies [55,68,69] provided personalized access to general health information. For example, Sanchez Bocanegra et al [15] targeted health-related videos and augmented them with trustworthy information from the United States National Library of Medicine (MedlinePlus) [110]. A total of 3 studies [52,72,74] related to health care forums focused on finding relevant threads. Cho et al [72] built “an autonomous agent that automatically responds to an unresolved user query by posting an automated response containing links to threads discussing similar medical problems.” However, 2 studies [71,73] helped patients to find similar patients. Jiang and Yang [71] investigated approaches for measuring user similarity in web-based health social websites, and Lima-Medina et al [73] built a virtual environment that facilitates contact among patients with cardiovascular problems. Both studies aim to help users seek informational and emotional support in a more efficient way. A total of 4 studies [41,75-77] helped patients to find appropriate doctors for a specific health problem, and 4 other studies [51,78-80] focused on finding nearby hospitals. A total of 2 studies [78,79] simply focused on the clinical preferences of the patients, whereas Krishnan et al [111] "provide health care recommendations that include Blood Donor recommendations and Hospital Specialization." Finally, Tabrizi et al [80] considered patient satisfaction as the primary feature of recommending hospitals to the user.

Specific Health Conditions

The last group of studies (9/73, 12%) focused on specific health conditions. However, the recommended items vary significantly. Torrent-Fontbona and Lopez Ibanez [81] have built a knowledge-based recommender system to assist diabetes patients in numerous cases, such as the estimated carbohydrate intake and past and future physical activity. Pustozerov et al [43] try to "reduce the carbohydrate content of the desired meal by reducing the amount of carbohydrate-rich products or by suggesting variants of products for replacement." Li and Kong [82] provided diabetes-related information, such as the need for a low-sodium lunch, targeted on American Indians through a mobile app. Other health conditions supported by recommender systems include depression and anxiety [83], mental disorders [45], and stress [34,54,84,85]. Both the mental disorder [45] and the depression and anxiety [83] HRSs recommend mobile apps. For example, the app MoveMe suggests exercises tailored to the user's mood. The HRS to alleviate stress includes recommending books to read [54] and meditative audios [85].

Recommender Techniques

Overview

The recommender techniques used varied greatly. Table 2 shows the distributions of these recommender techniques.

Table 2. Overview of the different recommender techniques used in the studies.

Main technique[a]	Study	Total studies, n (%)
Collaborative filtering	[59,69,76]	3 (4)
Content-based filtering	[15,32,54,63,72,86,87]	7 (10)
Knowledge-based filtering	[9,38,44,50,57,64,66,68,79,81,82,84,88-91]	16 (22)
Hybrid	[7,29,34,36,37,39-41,43,46-48,53,55,56,61,65,67,69,70,73,74,77,78,80,85,92-96,111]	32 (44)
Context-based techniques	[33,35,58,97]	4 (5)
Not specified	[45,83,98]	3 (4)
Comparison between techniques	[8,49,52,60,62,71,75,99]	8 (11)

[a] The papers are classified based on how the authors reported their techniques.

Recommender Techniques in Practice

The majority of HRSs (49/73, 67%) rely on knowledge-based techniques, either directly (17/49, 35%) or in a hybrid approach (32/49, 65%). Knowledge-based techniques are often used to incorporate additional information of patients into the recommendation process [112] and have been shown to improve the quality of recommendations while alleviating other drawbacks such as cold-start and sparsity issues [14]. Some studies use straightforward approaches, such as if-else reasoning based on domain knowledge [9,79,81,82,88,90,100]. Other studies use more complex algorithms such as particle swarm optimization [57], fuzzy logic [68], or reinforcement algorithms [44,84].

In total, 32 studies reported using a combination of recommender techniques and are classified as *hybrid recommender systems*. Different knowledge-based techniques are often combined. For example, Ali et al [56] used a combination of rule-based reasoning, case-based reasoning, and preference-based reasoning to recommend personalized physical activities according to the user's specific needs and personal interests. Asthana et al [46] combined the knowledge of a decision tree and demographic information to identify the health conditions. When health conditions are known, the system knows which measurements need to be monitored. A total of 7 studies used a *content-based technique* to recommend educational content [15,72,87], activities [32,86], reading materials [54], or nutritional advice [63].

Although collaborative filtering is a popular technique [113], it is not used frequently in the HRS domain. Marlin et al [59] used collaborative filtering to personalize future smoking cessation messages based on explicit feedback on past messages. This approach is used more often in combination with other techniques. A total of 2 studies [38,92] combined content-based techniques with collaborative filtering. Esteban et al [92], for instance, switched between content-based and collaborative approaches. The former approach is used for new physiotherapy exercises and the latter, when a new patient is registered or when previous recommendations to a patient are updated.

Context-Based Recommender Techniques

From an HRS perspective, context is described as an aggregate of various information that describes the setting in which an HRS is deployed, such as the location, the current activity, and the available time of the user. A total of 5 studies use contextual information to improve their recommendations but use a different technique; a prefilter uses contextual information to select or construct the most relevant data for generating recommendations. For example, in Narducci et al [75], the set of potentially similar patients was restricted to consultation requests in a specific medical area. Rist et al [33] applied a rule-based contextual prefiltering approach [114] to filter out inadequate recommendations, for example, "if it is dark outside, all outdoor activities, such as 'take a walk,' are filtered out" [33] before they are fed to the recommendation algorithm. However, a postfilter removes the recommended items *after* they are generated, such as filtering outdoor activities while it is raining. Casino et al [97] used a postfiltering technique by running the recommended items through a *real-time constraint checker*. Finally, contextual modeling, which was used by 2 studies [35,58], uses contextual information directly in the recommendation function as an explicit predictor of a user's rating for an item [114].

Location, agenda, and weather are examples of contextual information used by Lin et al [35] to promote the adoption of a healthy and active lifestyle. Cerón-Rios et al [58] used a decision tree to analyze user needs, health information, interests, time, location, and lifestyle to promote healthy habits. Casino et al [97] gathered contextual information through smart city sensor data to recommend healthier routes. Similarly, contextual information was acquired by Rist et al [33] using sensors embedded in the user's environment.

Comparisons

A total of 8 papers compared different recommender techniques to find the most optimal algorithm for a specific data set, end users, domain, and goal. Halder et al [52] used two well-known health forum data sets (PatientsLikeMe [115] and HealthBoards [116]) to compare 7 recommender techniques (among collaborative filtering and content-based filtering) and found that a hybrid approach scored best [52]. Another example is the study by Narducci et al [75], who compared four recommendation algorithms: cosine similarity as a baseline, collaborative filtering, their own *HealthNet* algorithm, and a hybrid of HealthNet and cosine similarity. They concluded that a prefiltering technique for similar patients in a specific medical area can drastically improve the recommendation accuracy [75]. The average and SD of the resulting ratings of the two collaborative techniques are compared with random recommendations by Li et al [60]. They show that a hybrid approach of a collaborative filter augmented with the calculated health level of the user performs better. In their nutrition-based meal recommender system, Yang et al [49] used item-wise and pairwise image comparisons in a two-step process. In conclusion, the 8 studies showed that recommendations can be improved when the benefits of multiple recommender techniques are combined in a hybrid solution [60] or contextual filters are applied [75].

Evaluation Approach

Overview

HRSs can be evaluated in multiple ways. In this study, we found two categories of HRS evaluations: (1) offline evaluations that use computational approaches to evaluate the HRS and (2) evaluations in which an end user is involved. Some studies used both, as shown in Multimedia Appendix 3.

Offline Evaluations

Of the total studies, 47% (34/73) do not involve users directly in their method of evaluation. The evaluation metrics also vary greatly, as many distinct metrics are reported in the included papers (Multimedia Appendix 3). Precision 53% (18/34), accuracy 38% (13/34), performance 35% (12/34), and recall 32% (11/34) were the most commonly used offline evaluation metrics. Recall has been used significantly more in recent papers, whereas accuracy also follows an upward trend. Moreover, performance was defined differently across studies. Torrent-Fontbona and Lopez Ibanez [81] compared the "amount of time in the glycaemic target range by reducing the time below the target" as performance. Cho et al [72] compared the precision and recall to report the performance. Clarke et al [84] calculated their own reward function to compare different approaches, and Lin et al [35] measured *system performance* as the number of messages sent in their in the wild study. Finally, Marlin et al [59] tested the predictive performance using a triple cross-validation procedure.

Other popular offline evaluation metrics are accuracy-related measurements, such as mean absolute (percentage) error, 18% (6/34); normalized discounted cumulative gain (nDCG), 18% (6/34); F_1 score, 15% (5/34); and root mean square error, 15% (5/34). The other metrics were measured inconsistently. For example, Casino et al [97] reported that they measure robustness but do not outline what they measure as robustness. However, they measured the mean absolute error. Torrent-Fontbona and Lopez Ibanez [81] defined robustness as the capability of the system to handle missing values. Effectiveness is also measured with different parameters, such as its ability to take the right classification decisions [75] or in terms of key opinion leaders' identification [41]. Finally, Li and Zaman [68] measured trust with a proxy: "evaluate the trustworthiness of a particular user in a health care social network based on factors such as role and reputation of the user in the social community" [68].

User Evaluations

Overview

Of the total papers, 53% (39/73) included participants in their HRS evaluation, with an average sample size of 59 (SD 84) participants (excluding the outlier of 8057 participants, as recruited in the study by Cheung et al [83]). On average, studies ran for more than 2 months (68, SD 56 days) and included all age ranges. There is a trend of increasing sample size and study duration over the years. However, only 17 studies reported the study duration; therefore, these trends were not significant. Surveys (12/39, 31%), user studies (10/39, 26%), and deployments in the wild (10/39, 26%) were the most used user evaluations. Only 6 studies used an RCT to evaluate their HRS. Finally, although all the included studies focused on HRSs and were dealing with sensitive data, only 12% (9/73) [9,34,42-45,73,83,95] reported ethical approval by a review board.

Surveys

No universal survey was found, as all the studies deployed a distinct survey. Ge et al [48] used the system usability scale and the framework of Knijnenburg et al [117] to explain the user experience of recommender systems. Esteban et al [95] designed their own survey with 10 questions to inquire about user experience. Cerón-Rios [58] relied on the ISO/IEC (International Organization of Standardization/International Electrotechnical Commission) 25000 standard to select 7 usability metrics to evaluate usability. Although most studies did not explicitly report the surveys used, user experience was a popular evaluation metric, as in the study by Wang et al [69]. Other metrics range from measuring user satisfaction [69,99] and perceived prediction accuracy [59] (with 4 self-composed questions). Nurbakova et al [98] combined data analytics with surveys to map their participants' psychological background, including orientations to happiness measured using the Peterson scale [118], personality traits using the Mini-International Personality Item Pool [119], and Fear of Missing Out based on the Przybylski scale [120].

Single-Session Evaluations (User Studies)

A total of 10 studies recruited users and asked them to perform certain tasks in a single session. Yang et al [49] performed a 60-person user study to assess its feasibility and effectiveness. Each participant was asked to rate meal recommendations relative to those made using a traditional survey-based approach. In a study by Gutiérrez et al [63], 15 users were asked to use the health augmented reality assistant and measure the qualities of the recommender system, users' behavioral intentions, perceived usefulness, and perceived ease of use. Jiang and Xu [77] performed 30 consultations and invited 10 evaluators majoring in medicine and information systems to obtain an average rating score and nDCG. Radha et al [8] used comparative questions to evaluate the feasibility. Moreover, Cheng et al [89] used 2 user studies to rank two *degrees of compromise* (DOC). A low DOC assigns more weight to the algorithm, and a high DOC assigns more weight to the user's health perspective. Recommendations with a lower DOC are more efficient for the user's health, but recommendations with a high DOC could convince users to believe that the recommended action is worth doing. Other approaches used are structured interviews [58], ranking [86,89], asking for unstructured feedback [40,88], and focus group discussions [87]. Finally, 3 studies [15,75,90] evaluated their system through a heuristic evaluation with expert users.

In the Wild

Only 2 studies tested their HRS *into the wild* recruited patients (people with a diagnosed health condition) in their evaluation. Yom-Tov et al [44] provided 27 sedentary patients with type 2 diabetes with a smartphone-based pedometer and a personal plan for physical activity. They assessed the effectiveness by calculating the amount of activity that the patient performed after the last message was sent. Lima-Medina et al [73] interviewed 45 patients with cardiovascular problems after a 6-month study period to measure (1) social management results, (2) health care plan results, and (3) recommendation results. Rist et al [33] performed an in-situ evaluation in an apartment of an older couple and used the data logs to describe the usage but augmented the data with a structured interview.

Yang et al [49] conducted a field study of 227 anonymous users that consisted of a training phase and a testing phase to assess the prediction accuracy. Buhl et al [99] created three user groups according to the recommender technique used and analyzed log data to compare the response rate, open email rate, and consecutive log-in rate. Similarly, Huang et al [76] compared the ratio of recommended doctors chosen and reserved by patients with the recommended doctors. Lin et al [35] asked 6 participants to use their HRSs for 5 weeks, measured system performance, studied user feedback to the recommendations, and concluded with an open-user interview. Finally, Ali et al [56] asked 10 volunteers to use their weight management systems for a couple of weeks. However, they do not focus on user-centric evaluation, as “only a prototype of the [...] platform is implemented.”

Rabbi et al [7] followed a single case with multiple baseline designs [121]. Single-case experiments achieve sufficient statistical power with a large number of repeated samples from a single individual. Moreover, Rabbi et al [7] argued that HRSs suit this requirement “since enough repeated samples can be collected with automated sensing or daily manual logging [121].” Participants were exposed to 2, 3, or 4 weeks of the control condition. The study ran for 7-9 weeks to compensate for the novelty effects. Food and exercise log data were used to measure changes in food calorie intake and calorie loss during exercise.

Randomized Controlled Trials

Only 6 studies followed an RCT approach. In the RCT by Bidargaddi et al [45], a large group of patients (n=192) and control group (n=195) were asked to use a web-based recommendation service for 4 weeks that recommended mental health and well-being mobile apps. Changes in well-being were measured using the Mental Health Continuum-Short Form [122]. The RCT by Sadasivam et al [42] enrolled 120 current smokers (n=74) and control group (n=46) as a follow-up to a previous RCT [123] that evaluated their portal to specifically evaluate the HRS algorithm. Message ratings were compared between the intervention and control groups.

Cheung et al [83] measured app loyalty through the number of weekly app sessions over a period of 16 weeks with 8057 users. In the study by Paredes et al [34], 120 participants had to use the HRS for at least 26 days. Self-reported stress assessment was performed before and after the intervention. Agapito et al [67] used an RCT with 40 participants to validate the sensitivity (true positive rate/[true positive rate+false negative rate]) and specificity (true negative rate/[true negative rate+false positive rate]) of the DIETOS HRS. Finally, Luo et al [93] performed a small clinical trial for more than 3 months (but did not report the number of participants). Their primary outcome measures included two standard clinical blood tests: fasting blood glucose and laboratory-measured glycated hemoglobin, before and after the intervention.

Interface

Overview

Only 47% (34/73) of the studies reported implementing a graphical user interface to communicate the recommended health items to the user. As illustrated in Table 3, 53% (18/34) use a mobile interface, usually through a mobile (web) app, whereas 36% (14/34) use a web interface to show the recommended items. Rist et al [33] built a kiosk into older adults' homes, as illustrated in Figure 3. Gutiérrez et al [63] used Microsoft HoloLens to project healthy food alternatives in augmented reality surrounding a physical object that the user holds, as shown in Figure 4.

Table 3. Distribution of the interfaces used among the different health recommender systems (n=34).

Interface	Study	Total studies, n (%)
Mobile	[7,34,35,40,44,48,56,58,66,69,77,78,82-84,86,88,97]	18 (53)
Web	[9,15,37,41,45,49,61,70,73,75,79,85,90,95]	14 (41)
Kiosk	[33]	1 (3)
HoloLens	[63]	1 (3)

Figure 3. Rist et al installed a kiosk in the home of older adults as a direct interface to their health recommender system.

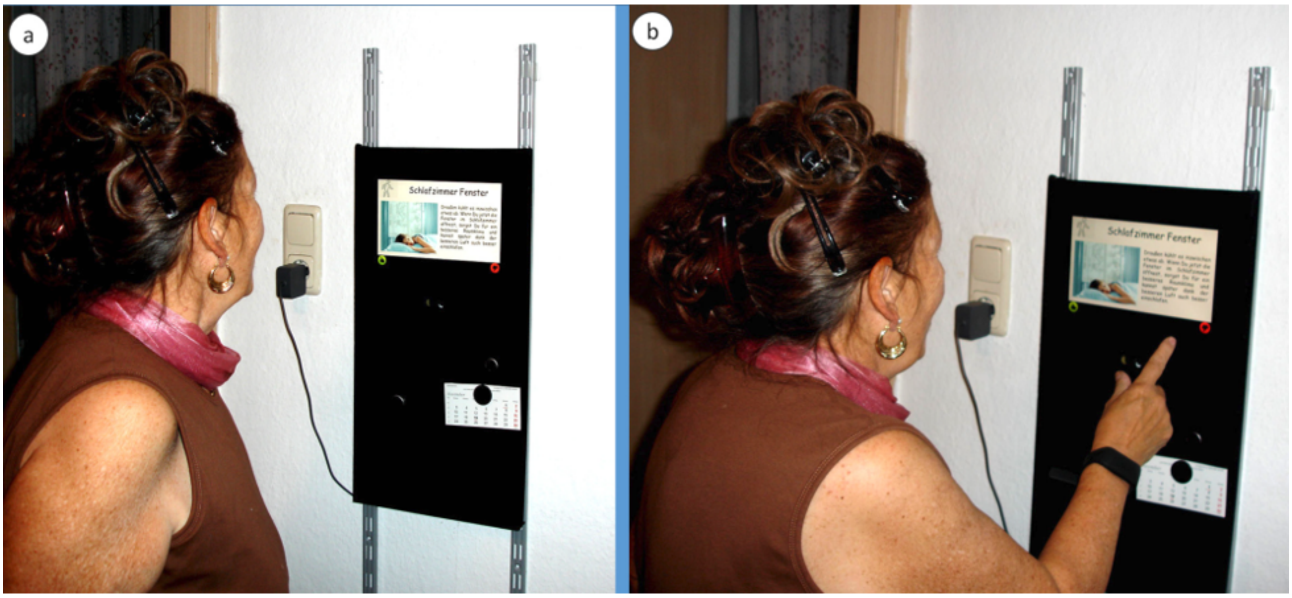


Figure 4. An example of the recommended healthy alternatives by Gutiérrez et al.

Visualization

A total of 7 studies [33,34,37,63,79,88,97] or approximately one-fourth of the studies with an interface included visualizations. However, the approach used was different for all studies, as shown in Table 4. Showing stars to show the relevance of a recommended item are only used by Casino et al [97] and Gutiérrez et al [63]. Wayman and Madhvanath [37] also used bar charts to visualize the progress toward a health goal. They visualize the healthy proportions, that is, what the user should eat. Somewhat more complex visualizations are used by Ho and Chen [88] who visualized the user's ECG zones. Paredes et al [34] presented an emotion graph as an input screen. Rist et al [33] visualized an example of how to perform the recommended activity.

Table 4. Distribution of the visualizations used among the different health recommender systems (n=7).

Visualization technique	Study	Total studies, n (%)
Bar charts	Wayman and Madhvanath [37] and Gutiérrez et al [63]	2 (29)
Heatmap	Ho and Chen [88]	1 (14)
Emotion graph	Paredes et al [34]	1 (14)
Visual example of action	Rist et al [33]	1 (14)
Map	Avila-Vazquez et al [79]	1 (14)
Star rating	Casino et al [97]	1 (14)

Transparency

In the study by Lage et al [87], participants expressed that:

> *they would like to have more control over recommendations received. In that sense, they suggested more information regarding the reasons why the recommendations are generated and more options to assess them.*

A total of 7 studies [7,37,41,45,63,66,82] explained the reasoning behind recommendations to end users at the user interface. Gutiérrez et al [63] provided recommendations for healthier food products and mentioned that the items (Figure 4) are based on the users' profile. Ueta et al [66] explained the relationship between the recommended dishes and a person's health conditions. For example, a person with acne can see the following text: “15 dishes that contained Pantothenic acid thought to be effective in acne a lot became a hit” [66]. Li and Kong [82] showed personalized recommended health actions in a message center. Color codes are used to differentiate between reminders, missed warnings, and recommendations. Rabbi et al [7] showed tailored motivational messages to explain why activities are recommended. For example, when the activity *walk near East Ave* is recommended, the app shows the additional message:

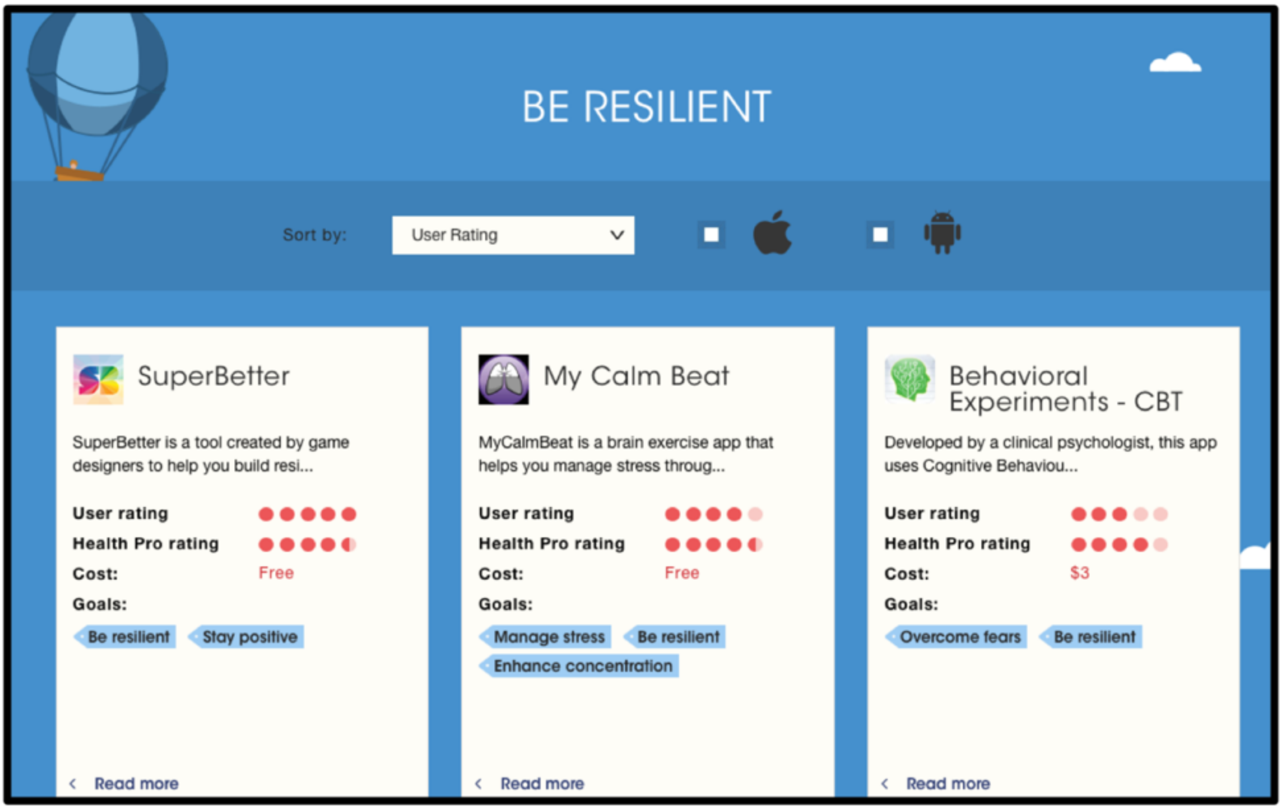
> *1082 walks in 240 days, 20 mins of walk everyday. Each walk nearly 4 min. Let us get 20 mins or more walk here today* [7]

Wayman and Madhvanath [37] first visualized the user's personal nutrition profile and used the lower part of the interface to explain why the item was recommended. They provided an illustrative example of spaghetti squash. The explanation shows that:

> *This product is high in Dietary_fiber, which you could consume more of. Try to get 3 servings a week* [37]

Guo et al [41] recommended doctors and showed a horizontal bar chart to visualize the user's values compared with the average values. Finally, Bidargaddi et al [45] visualized how the recommended app overlaps with the goal set by the users, as illustrated in Figure 5.

Figure 5. A screenshot from the health recommender system of Bidargaddi et al. Note the blue tags illustrating how each recommended app matches the users' goals.



BE RESILIENT

Sort by: User Rating

SuperBetter

SuperBetter is a tool created by game designers to help you build resi...

User rating ●●●●●
Health Pro rating ●●●●◐
Cost: Free
Goals:
Be resilient Stay positive

Read more

My Calm Beat

MyCalmBeat is a brain exercise app that helps you manage stress throug...

User rating ●●●●○
Health Pro rating ●●●●◐
Cost: Free
Goals:
Manage stress Be resilient Enhance concentration

Read more

Behavioral Experiments - CBT

Developed by a clinical psychologist, this app uses Cognitive Behaviou...

User rating ●●●○○
Health Pro rating ●●●●○
Cost: $3
Goals:
Overcome fears Be resilient

Read more

Discussion

Principal Findings

HRSs cover a multitude of subdomains, recommended items, implementation techniques, evaluation designs, and means of communicating the recommended items to the target user. In this systematic review, we clustered the recommended items into four groups: lifestyle, nutrition, general health care information, and specific health conditions. There is a clear trend toward HRSs that provide *well-being* recommendations but do not directly intervene in the user's medical status. For example, almost 70% (50/73; lifestyle and nutrition) focused on no strict medical recommendations. In the lifestyle group, physical activities (10/24, 42%) and advice on how to potentially change behavior (7/24, 29%) were recommended most often. In the nutrition group, these recommendations focused on nutritional advice (8/26, 31%), diets (7/26, 27%), and recipes (7/26, 27%). A similar trend was observed in the health care information group, where HRSs focused on guiding users to the appropriate environments such as hospitals (5/23, 22%) and medical professionals (4/23, 17%) or on helping users find qualitative information (5/23, 22%) on validated sources or from experiences by similar users and patients on health care forums (3/23, 13%). Thus, they only provide general information and do not intervene by recommending, for example, changing medication. Finally, when HRSs targeted specific health conditions, they recommended nonintervening actions, such as meditation sessions [84] or books to read [54].

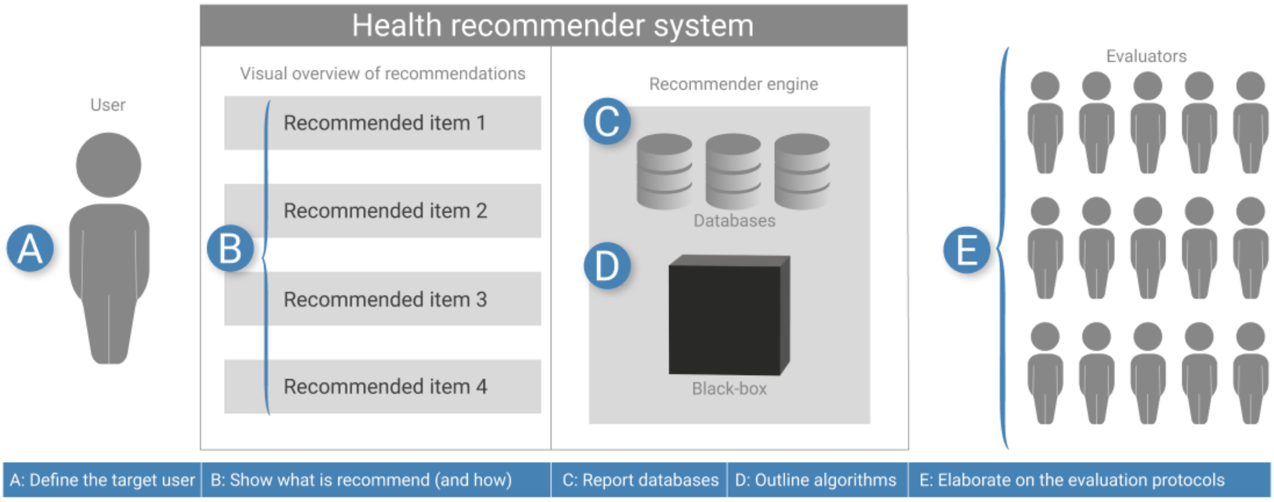
Although collaborative filtering is commonly the most used technique in other domains [124], here only 3 included studies reported the use of a collaborative filtering approach. Moreover, 43% (32/73) of the studies applied a hybrid approach, showing that HRS data sets might need special attention, which might also be the reason why all 73 studies used distinct data sets. In addition, the HRS evaluations varied greatly and were divided over evaluations where the end user was involved and evaluations that did not evolve users (offline evaluations). Only 47% (34/73) of the studies reported implementing a user interface to communicate recommendations to the user, despite the need to show the rationale of recommendations, as echoed by many researchers and practitioners [11]. Moreover, only 15% (7/47) included a (basic) visualization.

Unfortunately, this general lack of agreement on how to report HRSs might introduce researcher bias, as a researcher is currently completely unconstrained in defining what and how to measure the added value of an HRS. Therefore, further debate in the health recommender community is needed on how to define and measure the impact of HRSs. On the basis of our review and contribution to this discussion, we put forward a set of essential information that researchers should report in their studies.

Considerations for Practice

The previously discussed results have direct implications in practice and provide suggestions for future research. Figure 6 shows a reference frame of these requirements that can be used in future studies as a quality assessment tool.

Figure 6. A reference frame to report health recommender system studies. On the basis of the results of this study, we suggest that it should be clear what and how items are recommended (A), who the target user is (B), which data are used (C), and which recommender techniques are applied (D). Finally, the evaluation design should be reported in detail (E).



Define the Target User

As shown in this review, HRSs are used in a plethora of subdomains and each domain has its own experts. For example, in nutrition, the expert is most likely a dietician. However, the user of an HRS is usually a layperson without the knowledge of these domain experts, who often have different viewing preferences [125]. Furthermore, each user is unique. All individuals have idiosyncratic reasons for why they act, think, behave, and feel in a certain way at a specific stage of their life [126]. Not everybody is motivated by the same elements. Therefore, it is important to know the target user of the HRS. What is their previous knowledge, what are their goals, and what motivates them to act on a recommended item?

Show What Is Recommended (and How)

Researchers have become aware that *accuracy* is not sufficient to increase the effectiveness of a recommender system [127]. In recent years, research on human factors has gained attention. For example, He et al [11] surveyed 24 existing interactive recommender systems and compared their transparency, justification, controllability, and diversity. However, none of these 24 papers discussed HRSs. This indicates the gap between HRSs and recommender systems in other fields. Human factors have gained interest in the recommender community by “combining interactive visualization techniques with recommendation techniques to support transparency and controllability of the recommendation process” [11]. However, in this study, only 10% (7/73) explained the rationale of recommendations and only 10% (7/73) included a visualization to communicate the recommendations to the user. We do not argue that all HRSs should include a visualization or an explanation. However, researchers should pay attention to the delivery of these recommendations. Users need to understand, believe, and trust the recommended items before they can act on it.

To compare and assess HRSs, researchers should unambiguously report what the HRS is recommending. After all, typical recommender systems act like a *black box*, that is, they show suggestions without explaining the provenance of these recommendations [11]. Although this approach is suitable for typical e-commerce applications that involve little risk, transparency is a core requirement in higher risk application domains such as health [128]. Users need to understand why a recommendation is made, to assess its value and importance [12]. Moreover, health information can be cumbersome and not always easy to understand or situate within a specific health condition [129]. Users need to know whether the recommended item or action is based on a trusted source, tailored to their needs, and actionable [130].

Report the Data Set Used

All 73 studies used a distinct data set. Furthermore, some studies combine data from multiple databases, making it even more difficult to judge the quality of the data [35]. Nonetheless, most studies use self-generated data sets. This makes it difficult to compare and externally validate HRSs. Therefore, we argued that researchers should clarify the data used and potentially share whether these data are publicly available. However, in health data are often highly privacy sensitive and cannot be shared among researchers.

Outline the Recommender Techniques

The results show that there is no panacea for which recommender technique to use. The included studies differ from logic filters to traditional recommender techniques, such as collaborative filtering and content-based filtering to hybrid solutions and self-developed algorithms. However, with 44% (32/73), there is a strong trend toward the use of hybrid recommender techniques. The low number of collaborative filter techniques might be related to the fact that the evaluation sample sizes were also relatively low. Unfortunately, some studies have not fully disclosed the techniques used and only reported on the main algorithm used. It is remarkable that studies published in high-impact journals, such as studies by Bidargaddi et al [45] and Cheung et al [83], did not provide information on the recommender technique used. Nonetheless, disclosing the recommender technique allows other researchers not only to build on empirically tested technologies but also to verify whether key variables are included [29]. User data and behavior data can be identified to augment theory-based studies [29]. Researchers should prove that the algorithm is capable of recommending valid and trustworthy recommendations to the user based on their available data set.

Elaborate on the Evaluation Protocols

HRSs can be evaluated using different evaluation protocols. However, the protocol should be outlined mainly by the research goals of the authors. On the basis of the papers included in this study, we differentiate between the two approaches. In the first approach, the authors aim to influence their users' health, for example, by providing personalized diabetes guidelines [81] or prevention exercises for users with low back pain [95]. Therefore, the end user should always be involved in both the design and evaluation processes. However, only 8% (6/73) performed an RCT and 14% (10/73) deployed their HRS in the wild. This lack of user involvement has been noted previously by researchers and has been identified as a major challenge in the field [27,28]. Nonetheless, in other domains, such as job recommenders [131] or agriculture [132], user-centered design has been proposed as an important methodology in the design and development of tools used by end users, with the purpose of gaining trust and promoting technology acceptance, thereby increasing adoption with end users. Therefore, we recommend that researchers evaluate their HRSs with actual users. A potential model for a user-centric approach to recommender system evaluation is the user-centric framework proposed by Knijnenburg et al [117].

Research protocols need to be elaborated and approved by an ethical review board to prevent any impact on users. Authors should report how they informed their users and how they safeguarded the privacy of the users. This is in line with the modern journal and conference guidelines. For example, editorial policies of the *Journal of Medical Internet Research* state that "when reporting experiments on human subjects, authors should indicate IRB (Institutional Rese[a]rch Board, also known as REB) approval/exemption and whether the procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation" [133]. However, only 12% (9/73) reported their approval by an ethical review board. Acquiring review board approval will help the field mature and transition from small incremental studies to larger studies with representative users to make more reliable and valid findings.

In the second approach, the authors aim to design a *better* algorithm, where *better* is again defined by the authors. For example, the algorithm might perform faster, be more accurate, and be more efficient in computing power. Although the F_1 score, the mean absolute error, and nDCG are well defined and known within the recommender domain, other parameters are more ambiguous. For example, the performance or effectiveness can be assessed using different measurements. However, a health parameter can be monitored, such as the duration that a user remains within healthy ranges [81]. Furthermore, it could be a predictive parameter, such as an improved precision and recall as a proxy for performance [72]. Unfortunately, this difference makes it difficult to compare health recommendation algorithms. Furthermore, this inconsistency in measurement variables makes it infeasible to report in this systematic review which recommender techniques to use. Therefore, we argue that HRS algorithms should always be evaluated for other researchers to validate the results, if needed.

Limitations

This study has some limitations that affect its contribution. Although an extensive scope search was conducted in scientific databases and most relevant health care informatic journals, some relevant literature in other domains might have been excluded. The keywords used in the search string could have impacted the results. First, we did not include domain-specific constructs of health, such as asthma, pregnancy, and iron deficiency. Many studies may implicitly report healthy computer-generated recommendations when they research the impact of a new intervention. In these studies, however, building an HRS is often not their goal and, therefore, was excluded from this study. Second, we searched for papers that reported studying an HRS; nonincluded studies might have built an HRS but did not report it as such. Considering our RQs, we deemed it important that authors explicitly reported their work as a recommender system. To conclude, in this study, we provide a large cross-domain overview of health recommender techniques targeted to laypersons and deliver a set of recommendations that could help the field of HRS mature.

Conclusions

This study presents a comprehensive report on the use of HRS across domains. We have discussed the different subdomains HRS applied in, the different recommender techniques used, the different manners in which they are evaluated, and finally, how they present the recommendations to the user. On the basis of this analysis, we have provided research guidelines toward a consistent reporting of HRSs. We found that although most applications are intended to improve users' well-being, there is a significant opportunity for HRSs to inform and guide users' health actions. Although many of the studies present a lack of a user-centered evaluation approach, some studies performed full-scale RCT evaluations or elaborated in the wild studies to validate their HRS, showing the field of HRS is slowly maturing. On the basis of this study, we argue that it should always be clear what the HRS is recommending and to whom these recommendations are for. Graphical assets should be added to show how recommendations are presented to users. Authors should also report which data sets and algorithms were used to calculate the recommendations. Finally, detailed evaluation protocols should be reported.

We conclude that the results motivate the creation of richer applications in future design and development of HRSs. The field is maturing, and interesting opportunities are being created to inform and guide health actions.

Acknowledgments

This work was part of the research project PANACEA Gaming Platform with project HBC.2016.0177, which was financed by Flanders Innovation & Entrepreneurship and research project IMPERIUM with research grant G0A3319N from the Research Foundation-Flanders (FWO) and the Slovenian Research Agency grant ARRS-N2-0101. Project partners were BeWell Innovations and the University Hospital of Antwerp.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Coded data set of all included papers.
[XLS File (Microsoft Excel File), 50 KB - jmir_v23i6e18035_app1.xls]

Multimedia Appendix 2
Overview of recommended items by 73 studies.
[PNG File , 323 KB - jmir_v23i6e18035_app2.png]

Multimedia Appendix 3
Overview of evaluation approaches.
[PNG File , 375 KB - jmir_v23i6e18035_app3.png]

References

1. Stevens G, Mascarenhas M, Mathers C. Global health risks: progress and challenges. Bull World Health Organ 2009 Sep;87(9):646 [FREE Full text] [doi: 10.2471/blt.09.070565] [Medline: 19784438]
2. de Choudhury M, Sharma S, Kiciman E. Characterizing Dietary Choices, Nutrition, and Language in Food Deserts via Social Media. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing New York, NY. New York, United States: Association for Computing Machinery; 2016 Mar 2 Presented at: CSCW'16; Feb 27–March 2, 2016; San Francisco p. 1157-1170. [doi: 10.1145/2818048.2819956]
3. Glanz K, Rimer B, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco: John Wiley & Sons; 2008.
4. Woolf SH. The power of prevention and what it requires. J Am Med Assoc 2008 May 28;299(20):2437-2439. [doi: 10.1001/jama.299.20.2437] [Medline: 18505953]
5. Elsweiler D, Ludwig B, Said A, Schaefer H, Trattner C. Engendering Health with Recommender Systems. In: Proceedings of the 10th ACM Conference on Recommender Systems. New York, United States: Association for Computing Machinery; 2016 Presented at: RecSys'16; September 15 - September 19, 2016; New York, NY p. -. [doi: 10.1145/2959100.2959203]
6. Yürüten O. Recommender Systems for Healthy Behavior Change Internet. École polytechnique fédérale de Lausanne. 2017. URL: https://infoscience.epfl.ch/record/231155/files/EPFL_TH7973.pdf [accessed 2021-06-04]
7. Rabbi M, Aung MH, Zhang M, Choudhury T. MyBehavior: Automatic Personalized Health Feedback from User Behaviors Preferences Using Smartphones. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive Ubiquitous Computing New York, NY. New York, United States: Association for Computing Machinery; 2015 Presented at: UbiComp'15; September 7-11, 2015; Osaka p. A. [doi: 10.1145/2750858.2805840]
8. Radha M, Willemsen MC, Boerhof M, IJsselsteijn WA. Lifestyle Recommendations for Hypertension through Rasch-based Feasibility Modeling. In: Proceedings of the 2016 Conference on User Modeling Adaptation Personalization. New York, United States: Association for Computing Machinery; 2016 Presented at: UMAP'16; July 13–17, 2016; Halifax, Nova Scotia, Canada p. 239-247. [doi: 10.1145/2930238.2930251]
9. Hidalgo JI, Maqueda E, Risco-Martín JL, Cuesta-Infante A, Colmenar JM, Nobel J. glUCModel: a monitoring and modeling system for chronic diseases applied to diabetes. J Biomed Inform 2014 Apr;48:183-192 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.015] [Medline: 24407050]
10. Gro¨nvall E, Verdezoto N, Bagalkot N, Sokoler T. Concordance: A Critical Participatory Alternative in Healthcare IT. In: Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives. 2015 Oct 5 Presented at: CA'15; August 17-21, 2015; Aarhus, Denmark p. 21-24. [doi: 10.7146/aahcc.v1i1.21315]
11. He C, Parra D, Verbert K. Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities. Expert Syst Appl 2016 Sep;56:9-27. [doi: 10.1016/j.eswa.2016.02.013]
12. Ricci F, Rokach L, Shapira B. Recommender Systems Handbook. USA: Springer US; 2015:1003.
13. Kobsa A, Nejdl W. Hybrid web recommender systems. In: Brusilovsky P, Kobsa A, Nejdl W, editors. The Adaptive Web Methods and Strategies of Web Personalization. Berlin: Springer Berlin Heidelberg; 2007:377-480.

14. Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 2005 Jun;17(6):734-749. [doi: 10.1109/tkde.2005.99]
15. Bocanegra CL, Sevillano 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]
16. Wu L, Shah S, Choi S, Tiwari M, Posse C. The Browsemaps: Collaborative Filtering at LinkedIn. In: Proceedings of the 6th Workshop on Recommender Systems and the Social Web. 2014 Presented at: RecSys'14; October 6-10, 2014; Foster City, CA, USA URL: http://ceur-ws.org/Vol-1271/Paper3.pdf
17. Verbert K, Duval E, Lindstaedt SN, Gillet D. Context-aware recommender systems. J Univers Comput Sci 2010;16(16):2175-2178. [doi: 10.1007/978-0-387-85820-3_7]
18. Verbert K, Manouselis N, Ochoa X, Wolpers M, Drachsler H, Bosnic I, et al. Context-aware recommender systems for learning: a survey and future challenges. IEEE Trans Learn Technol 2012 Oct;5(4):318-335. [doi: 10.1109/tlt.2012.11]
19. Mahmood T, Ricci F. Improving Recommender Systems With Adaptive Conversational Strategies. In: Proceedings of the 20th ACM Conference on Hypertext Hypermedia. USA: Association for Computing Machinery; 2009 Presented at: ACM'09; June 29–July 1, 2009; Torino, Italy p. 73-85. [doi: 10.1145/1557914.1557930]
20. Resnick P, Varian HR. Recommender systems. Commun ACM 1997 Mar;40(3):56-58. [doi: 10.1145/245108.245121]
21. Wiesner M, Pfeifer D. Health recommender systems: concepts, requirements, technical basics and challenges. Int J Environ Res Public Health 2014 Mar 3;11(3):2580-2607 [FREE Full text] [doi: 10.3390/ijerph110302580] [Medline: 24595212]
22. Sezgin E, Özkan S. A Systematic Literature Review on Health Recommender Systems. In: E-Health and Bioengineering Conference (EHB).: IEEE; 2013 Presented at: EHB'13; November 21-23, 2013; Iasi, Romania p. 1. [doi: 10.1109/ehb.2013.6707249]
23. Valdez AC, Ziefle M, Verbert K, Felfernig A, Holzinger A. Recommender systems for health informatics: state-of-the-art and future perspectives. In: Holzinger A, editor. Machine Learning for Health Informatics. Lecture Notes in Computer Science. Cham, Switzerland: Springer International Publishing; 2016:391-414.
24. Kamran M, Javed A. A survey of recommender systems and their application in healthcare. Tech J 2015;20(4):111-119 [FREE Full text]
25. Afolabi AO, Toivanen P, Haataja K, Mykkänen J. Systematic literature review on empirical results and practical implementations of healthcare recommender systems: lessons learned and a novel proposal. Int J Healthcare Inform Syst Inf 2015;10(4):1-21 [FREE Full text] [doi: 10.4018/IJHISI.2015100101]
26. Ferretto LR, Cervi CR, de Marchi AC. Recommender systems in mobile apps for health a systematic review. In: 12th Iberian Conference on Information Systems and Technologies (CISTI).: IEEE; 2017 Presented at: CISTI'17; June 21-24, 2017; Lisbon, Portugal p. 1-6. [doi: 10.23919/cisti.2017.7975743]
27. Hors-Fraile S, Rivera-Romero O, Schneider F, Fernandez-Luque L, Luna-Perejon F, Civit-Balcells A, et al. Analyzing recommender systems for health promotion using a multidisciplinary taxonomy: a scoping review. Int J Med Inform 2018 Jun;114:143-155. [doi: 10.1016/j.ijmedinf.2017.12.018] [Medline: 29331276]
28. Schäfer H, Hors-Fraile S, Karumur RP, Valdez AC, Said A, Torkamaan H, et al. Towards Health (Aware) Recommender Systems. In: Proceedings of the 2017 International Conference on Digital Health. USA: ACM; 2017 Presented at: DH'17; July 2-5, 2017; London p. 157-161. [doi: 10.1145/3079452.3079499]
29. Sadasivam RS, Cutrona SL, Kinney RL, Marlin BM, Mazor KM, Lemon SC, et al. Collective-intelligence recommender systems: advancing computer tailoring for health behavior change into the 21st century. J Med Internet Res 2016 Mar 7;18(3):e42 [FREE Full text] [doi: 10.2196/jmir.4448] [Medline: 26952574]
30. Cappella JN, Yang S, Lee S. Constructing recommendation systems for effective health messages using content, collaborative, and hybrid algorithms. Ann Am Acad Pol Soc Sci 2015 Apr 9;659(1):290-306. [doi: 10.1177/0002716215570573]
31. Lika B, Kolomvatsos K, Hadjiefthymiades S. Facing the cold start problem in recommender systems. Expert Syst Appl 2014 Mar;41(4):2065-2073. [doi: 10.1016/j.eswa.2013.09.005]
32. Kumar G, Jerbi H, Gurrin C, O'Mahony M. Towards Activity Recommendation from Lifelogs. In: Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services. 2014 Presented at: CIWAS'14; December 4-6 2014; Hanoi, Viet Nam. [doi: 10.1145/2684200.2684298]
33. Rist T, Seiderer A, Hammer S, Mayr M, André E. CARE - Extending a Digital Picture Frame with a Recommender Mode to Enhance Well-Being of Elderly People. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. 2015 Presented at: IEEE'115; May 20-23, 2015; Istanbul, Turkey. [doi: 10.4108/icst.pervasivehealth.2015.259255]
34. Paredes P, Gilad-Bachrach R, Czerwinski M, Roseway A, Rowan K, Hernandez J. PopTherapy: Coping With Stress Through Pop-culture. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare. 2014 Presented at: PCTH'14; May 20-23, 2014; Oldenburg, Germany. [doi: 10.4108/icst.pervasivehealth.2014.255070]
35. Lin Y, Jessurun J, de Vries B, Timmermans H. Motivate: towards Context-Aware Recommendation Mobile System for Healthy Living. In: 5th International ICST Conference on Pervasive Computing Technologies for Healthcare. 2011 Presented at: PCTH'11; May 23-26, 2011; Dublin, Republic of Ireland. [doi: 10.4108/icst.pervasivehealth.2011.246030]

36. Elsweiler D, Harvey M. Towards Automatic Meal Plan Recommendations for Balanced Nutrition. In: Proceedings of the 9th ACM Conference on Recommender Systems. 2015 Presented at: ACM'15; September 16–20, 2015; Vienna, Austria. [doi: 10.1145/2792838.2799665]
37. Wayman E, Madhvanath S. Nudging Grocery Shoppers to Make Healthier Choices. In: Proceedings of the 9th ACM Conference on Recommender Systems. 2015 Presented at: ACM'15; September 16-20, 2015; Vienna, Austria. [doi: 10.1145/2792838.2799669]
38. Subagdja B, Tan AH. Coordinated Persuasion with Dynamic Group Formation for Collaborative Elderly Care. In: 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2015 Presented at: WI-IAT'15; December 6-9, 2015; Singapore. [doi: 10.1109/wi-iat.2015.77]
39. Lafta R, Zhang J, Tao X, Li Y, Tseng VS. An Intelligent Recommender System Based on Short-Term Risk Prediction for Heart Disease Patients. In: 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2015 Presented at: WI-IAT'15; December 6-9, 2015; Singapore. [doi: 10.1109/wi-iat.2015.47]
40. Takama Y, Sasaki W, Okumura T, Yu C, Chen L, Ishikawa H. Walking Route Recommendation System for Taking a Walk as Health Promotion. In: 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2015 Presented at: WI-IAT'15; December 6-9, 2015; Singapore. [doi: 10.1109/wi-iat.2015.218]
41. Guo L, Jin B, Yao C, Yang H, Huang D, Wang F. Which doctor to trust: a recommender system for identifying the right doctors. J Med Internet Res 2016 Jul 7;18(7):e186 [FREE Full text] [doi: 10.2196/jmir.6015] [Medline: 27390219]
42. Sadasivam RS, Borglund EM, Adams R, Marlin BM, Houston TK. Impact of a collective intelligence tailored messaging system on smoking cessation: the perspect randomized experiment. J Med Internet Res 2016 Nov 8;18(11):e285 [FREE Full text] [doi: 10.2196/jmir.6465] [Medline: 27826134]
43. Pustozerov E, Popova P, Tkachuk A, Bolotko Y, Yuldashev Z, Grineva E. Development and evaluation of a mobile personalized blood glucose prediction system for patients with gestational diabetes mellitus. JMIR Mhealth Uhealth 2018 Jan 9;6(1):e6 [FREE Full text] [doi: 10.2196/mhealth.9236] [Medline: 29317385]
44. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. J Med Internet Res 2017 Oct 10;19(10):e338 [FREE Full text] [doi: 10.2196/jmir.7994] [Medline: 29017988]
45. Bidargaddi N, Musiat P, Winsall M, Vogl G, Blake V, Quinn S, et al. Efficacy of a web-based guided recommendation service for a curated list of readily available mental health and well-being mobile apps for young people: randomized controlled trial. J Med Internet Res 2017 May 12;19(5):e141 [FREE Full text] [doi: 10.2196/jmir.6775] [Medline: 28500020]
46. Asthana S, Megahed A, Strong R. A Recommendation System for Proactive Health Monitoring Using IoT and Wearable Technologies. In: 2017 IEEE International Conference on AI & Mobile Services. 2017 Presented at: AIMS'17; 25-30 June, 2017; Honolulu, HI, USA. [doi: 10.1109/aims.2017.11]
47. Ng YK, Jin M. Personalized Recipe Recommendations for Toddlers Based on Nutrient Intake and Food Preferences. In: Proceedings of the 9th International Conference on Management of Digital EcoSystems. 2017 Presented at: CMDE'17; November 7–10, 2017; Bangkok, Thailand. [doi: 10.1145/3167020.3167057]
48. Ge M, Elahi M, Fernaández-Tobías I, Ricci F, Massimo D. Using Tags and Latent Factors in a Food Recommender System. In: Proceedings of the 5th International Conference on Digital Health 2015. 2015 Presented at: CDH'15; May 18–20, 2015; Florence, Italy. [doi: 10.1145/2750511.2750528]
49. Yang L, Hsieh C, Yang H, Pollak J, Dell N, Belongie S, et al. Yum-me: a personalized nutrient-based meal recommender system. ACM Trans Inf Syst 2017 Aug;36(1):1-31 [FREE Full text] [doi: 10.1145/3072614] [Medline: 30464375]
50. Pawar KR, Ghorpade T, Shedge R. Constraint Based Recipe Recommendation Using Forward Checking Algorithm. In: International Conference on Advances in Computing, Communications and Informatics. 2016 Presented at: ICACCI'16; Spetember 21-24, 2016; Jaipur, India. [doi: 10.1109/icacci.2016.7732256]
51. Krishna S, Venkatesh P, Swagath S, Valliyammai C. A Trust Enhanced Recommender System for Medicare Applications. In: Sixth International Conference on Advanced Computing. 2014 Presented at: ICoAC'14; December 17-19, 2014; Chennai, India. [doi: 10.1109/icoac.2014.7229734]
52. Halder K, Kan MY, Sugiyama K. Health Forum Thread Recommendation Using an Interest Aware Topic Model. In: Proceedings of the 2017 ACM on Conference on Information Knowledge Management. 2017 Presented at: ACM'17; November 6–10, 2017; Singapore. [doi: 10.1145/3132847.3132946]
53. Farrell RG, Danis CM, Ramakrishnan S, Kellogg WA. Intrapersonal Retrospective Recommendation: Lifestyle Change Recommendations Using Stable Patterns of Personal Behavior. In: Proceedings of the First International Workshop on Recommendation Technologies for Lifestyle Change. 2012 Presented at: LIFESTYLE'12; September 13, 2012; Dublin, Ireland.
54. Xin Y, Chen Y, Jin L, Cai Y, Feng L. TeenRead: An Adolescents Reading Recommendation System Towards Online Bibliotherapy. In: IEEE International Congress on Big Data. 2017 Presented at: BigData Congress'17; June 25-30, 2017; Honolulu, HI, USA. [doi: 10.1109/bigdatacongress.2017.63]
55. Zaman N, Li J. Semantics-Enhanced Recommendation System for Social Healthcare. In: 28th International Conference on Advanced Information Networking and Applications. 2014 Presented at: IEEE'14; May 13–16, 2014; Victoria, Canada. [doi: 10.1109/aina.2014.93]

56. Ali R, Afzal M, Hussain M, Ali M, Siddiqi MH, Lee S, et al. Multimodal hybrid reasoning methodology for personalized wellbeing services. Comput Biol Med 2016 Feb 1;69:10-28. [doi: 10.1016/j.compbiomed.2015.11.013] [Medline: 26705863]
57. Pop CB, Chifu VR, Salomie I, Cozac A, Mesaros I. Particle Swarm Optimization-based Method for Generating Healthy Lifestyle Recommendations. In: 9th International Conference on Intelligent Computer Communication and Processing. 2013 Presented at: ICCP'13; September 5-7, 2013; Cluj-Napoca, Romania. [doi: 10.1109/iccp.2013.6646074]
58. Cerón-Rios G, López DM, Blobel B. Architecture and user-context models of cocare: a context-aware mobile recommender system for health promotion. Stud Health Technol Inform 2017;237:140-147. [Medline: 28479557]
59. Marlin BM, Adams RJ, Sadasivam R, Houston TK. Towards collaborative filtering recommender systems for tailored health communications. AMIA Annu Symp Proc 2013;2013:1600-1607 [FREE Full text] [Medline: 24551430]
60. Li H, Zhang Q, Lu K. Integrating Mobile Sensing and Social Network for Personalized Health-care Application. In: Proceedings of the 30th Annual ACM Symposium on Applied Computing. 2015 Presented at: ACM'15; April 13-17, 2015; Salamanca, Spain. [doi: 10.1145/2695664.2695767]
61. Ntalaperas D, Bothos E, Perakis K, Magoutas B, Mentzas G. An Intelligent System for Personalized Nutritional Recommendations in Restaurants. In: Proceedings of the 19th Panhellenic Conference on Informatics. 2015 Presented at: PCI'15; October 1-3, 2015; Athens, Greece. [doi: 10.1145/2801948.2801997]
62. Trattner C, Elsweiler D. Investigating the Healthiness of Internet-Sourced Recipes: Implications for Meal Planning and Recommender Systems. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; April 3–7, 2017; Perth, Australia. [doi: 10.1145/3038912.3052573]
63. Gutiérrez F, Cardoso B, Verbert K. PHARA: A Personal Health Augmented Reality Assistant to Support Decision-making at Grocery Stores. In: Proceedings of the International Workshop on Health Recommender Systems co-located with ACM. 2017 Presented at: RecSys'17; August 1-4, 2017; Como, Italy. [doi: 10.1145/3236112.3236161]
64. Lo YW, Zhao Q, Chen RC. Automatic Generation and Recommendation of Recipes Based on Outlier Analysis. In: 7th International Conference on Awareness Science and Technology. 2015 Presented at: iCAST'15; May 6-9, 2015; Qinhuangdao, China. [doi: 10.1109/icawst.2015.7314050]
65. Ooi C, Iiba C, Takano C. Ingredient Substitute Recommendation for Allergy-safe Cooking Based on Food Context. In: Pacific Rim Conference on Communications, Computers and Signal Processing. 2015 Presented at: PACRIM'15; August 24-26, 2015; Victoria, BC, Canada. [doi: 10.1109/pacrim.2015.7334878]
66. Ueta T, Iwakami M, Ito T. Implementation of a Goal-Oriented Recipe Recommendation System Providing Nutrition Information. In: International Conference on Technologies and Applications of Artificial Intelligence. 2011 Presented at: TAAI'11; November 11-13, 2011; Chung Li, Taiwan. [doi: 10.1109/taai.2011.39]
67. Agapito G, Simeoni M, Calabrese B, Caré I, Lamprinoudi T, Guzzi PH, et al. DIETOS: a dietary recommender system for chronic diseases monitoring and management. Comput Methods Programs Biomed 2018 Jan;153:93-104. [doi: 10.1016/j.cmpb.2017.10.014] [Medline: 29157465]
68. Li J, Zaman N. Personalized Healthcare Recommender Based on Social Media. In: Proceedings of the 2014 IEEE 28th International Conference on Advanced Information Networking and Applications. 2014 Presented at: AINA'14; May 13-16, 2014; Victoria, BC, Canada. [doi: 10.1109/aina.2014.120]
69. Wang S, Chen YL, Kuo AM, Chen H, Shiu YS. Design and evaluation of a cloud-based mobile health information recommendation system on wireless sensor networks. Comput Electr Engg 2016 Jan;49:221-235. [doi: 10.1016/j.compeleceng.2015.07.017]
70. Wang CS, Li CY. Integrated Baby-care Recommender Platform Based on Hybrid Commonsense Reasoning and Case-based Reasoning Algorithms. In: The 6th International Conference on Networked Computing and Advanced Information Management. 2010 Presented at: NCAIM'10; August 16-18, 2010; Seoul, Korea (South).
71. Jiang L, Yang CC. User recommendation in healthcare social media by assessing user similarity in heterogeneous network. Artif Intell Med 2017 Sep;81:63-77. [doi: 10.1016/j.artmed.2017.03.002] [Medline: 28325605]
72. Cho JH, Sondhi P, Zhai C, Schatz BR. Resolving Healthcare Forum Posts via Similar Thread Retrieval. In: Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, Health Informatics. 2014 Presented at: ACM'14; September 20–23, 2014; Newport Beach, CA, USA. [doi: 10.1145/2649387.2649399]
73. Lima-Medina E, Loques O, Mesquita C. 'Minha Saude' a Healthcare Social Network for Patients With Cardiovascular Diseases. In: 3nd International Conference on Serious Games and Applications for Health (SeGAH). 2014 Presented at: SeGAH'14; May 14-16, 2014; Rio de Janeiro, Brazil. [doi: 10.1109/segah.2014.7067070]
74. Wang J, Man C, Zhao Y, Wang F. An Answer Recommendation Algorithm for Medical Community Question Answering Systems. In: IEEE International Conference on Service Operations and Logistics, and Informatics. 2016 Presented at: SOLI'16; July 10-12, 2016; Beijing, China. [doi: 10.1109/soli.2016.7551676]
75. Narducci F, Lops P, Semeraro G. Power to the patients: the HealthNetsocial network. Information Systems 2017 Nov;71:111-122. [doi: 10.1016/j.is.2017.07.005]
76. Huang Y, Liu P, Pan Q, Lin J. A Doctor Recommendation Algorithm Based on Doctor Performances and Patient Preferences. In: International Conference on Wavelet Active Media Technology and Information Processing. 2012 Presented at: ICWAMTIP'12; December 17-19, 2012; Chengdu, China. [doi: 10.1109/icwamtip.2012.6413447]

77. Jiang H, Xu W. How to Find Your Appropriate Doctor: an Integrated Recommendation Framework in Big Data Context. In: IEEE Symposium on Computational Intelligence in Healthcare and e-health. 2014 Presented at: CICARE'14; December 9-12, 2014; Orlando, FL, USA. [doi: 10.1109/cicare.2014.7007848]
78. Chen TT, Chiu M. Mining the preferences of patients for ubiquitous clinic recommendation. Health Care Manag Sci 2020 Jun 6;23(2):173-184. [doi: 10.1007/s10729-018-9441-y] [Medline: 29511976]
79. Ávila-Vázquez D, Lopez-Martinez M, Maya D, Olvera X, Guzmán G, Torres M, et al. Geospatial Recommender System for the Location of Health Services. In: 2014 9th Iberian Conference on Information Systems and Technologies. 2014 Presented at: CISTI'14; June 18-21, 2014; Barcelona, Spain. [doi: 10.1109/cisti.2014.6877023]
80. Tabrizi TS, Khoie MR, Sahebkar E, Rahimi S, Marhamati N. Towards a Patient Satisfaction Based Hospital Recommendation System. In: 2016 International Joint Conference on Neural Networks. 2016 Presented at: IJCNN'16; July 24-29, 2016; Vancouver, BC, Canada. [doi: 10.1109/ijcnn.2016.7727190]
81. Torrent-Fontbona F, Lopez B. Personalized adaptive CBR bolus recommender system for type 1 diabetes. IEEE J Biomed Health Inform 2019 Jan;23(1):387-394. [doi: 10.1109/JBHI.2018.2813424] [Medline: 29994082]
82. Li J, Kong J. Cell Phone-based Diabetes Self-management and Social Networking System for American Indians. In: IEEE 18th International Conference on e-Health Networking, Applications and Services. 2016 Presented at: Healthcom'16; September 14-16, 2016; Munich, Germany. [doi: 10.1109/healthcom.2016.7749456]
83. Cheung K, Ling W, Karr CJ, Weingardt K, Schueller SM, Mohr DC. Evaluation of a recommender app for apps for the treatment of depression and anxiety: an analysis of longitudinal user engagement. J Am Med Inform Assoc 2018 Aug 1;25(8):955-962 [FREE Full text] [doi: 10.1093/jamia/ocy023] [Medline: 29659857]
84. Clarke S, Jaimes LG, Labrador MA. Mstress: a Mobile Recommender System for Just-in-time Interventions for Stress. In: 14th IEEE Annual Consumer Communications & Networking Conference. 2017 Presented at: CCNC'17; January 8-11, 2017; Las Vegas, NV, USA. [doi: 10.1109/ccnc.2017.8015367]
85. Zaini N, Latip MF, Omar H, Mazalan L, Norhazman H. Online Personalized Audio Therapy Recommender Based on Community Ratings. In: International Symposium on Computer Applications and Industrial Electronics. 2012 Presented at: ISCAIE'12; December 3-4, 2012; Kota Kinabalu, Malaysia. [doi: 10.1109/iscaie.2012.6482120]
86. Bravo-Torres JF, Ordoñez-Ordoñez JO, Gallegos-Segovia PL, Vintimilla-Tapia PE, López-Nores M, Blanco-Fernández Y. A Context-aware Platform for Comprehensive Care of Elderly People: Proposed Architecture. In: CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies. 2017 Presented at: CHILECON'17; October 18-20, 2017; Pucon, Chile. [doi: 10.1109/chilecon.2017.8229507]
87. Lage R, Durao F, Dolog P, Stewart A. Applicability of Recommender Systems to Medical Surveillance Systems. In: Proceedings of the Second International Workshop on Web Science and Information Exchange in the Medical Web. 2011 Presented at: WSIE'11; October 28, 2011; Glasgow, Scotland, UK. [doi: 10.1145/2064741.2064743]
88. Ho TC, Chen X. ExerTrek: a Portable Handheld Exercise Monitoring, Tracking and Recommendation System. In: 11th International Conference on e-Health Networking, Applications and Services. 2009 Presented at: Healthcom'09; December 16-18, 2009; Sydney, NSW, Australia. [doi: 10.1109/health.2009.5406194]
89. Cheng CY, Qian X, Tseng SH, Fu LC. Recommendation Dialogue System Through Pragmatic Argumentation. In: 26th IEEE International Symposium on Robot and Human Interactive Communication. 2017 Presented at: RO-MAN'17; August 29-30, 2017; Lisbon, Portugal. [doi: 10.1109/roman.2017.8172323]
90. Wongpun S, Guha S. Elderly Care Recommendation System for Informal Caregivers Using Case-based Reasoning. In: 2nd Advanced Information Technology, Electronic and Automation Control Conference. 2017 Presented at: IAEAC'17; March 25-26, 2017; Chongqing, China. [doi: 10.1109/iaeac.2017.8054075]
91. Gai K, Qiu M, Jayaraman S, Tao L. Ontology-Based Knowledge Representation for Secure Self-Diagnosis in Patient-Centered Teleheath with Cloud Systems. In: 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing. 2015 Presented at: IEEE'15; November 3-5, 2015; New York, NY, USA. [doi: 10.1109/cscloud.2015.72]
92. Phanich M, Pholkul P, Phimoltares S. Food Recommendation System Using Clustering Analysis for Diabetic Patients. In: International Conference on Information Science and Applications. 2010 Presented at: CISA'10; April 21-23, 2010; Seoul, Korea (South). [doi: 10.1109/icisa.2010.5480416]
93. Luo Y, Ling C, Schuurman J, Petrella R. GlucoGuide: An Intelligent Type-2 Diabetes Solution Using Data Mining and Mobile Computing. In: IEEE International Conference on Data Mining Workshop.: IEEE; 2014 Presented at: IEEE'14; December 14, 2014; Shenzhen, China. [doi: 10.1109/icdmw.2014.177]
94. Mustaqeem A, Anwar SM, Khan AR, Majid M. A statistical analysis based recommender model for heart disease patients. Int J Med Inform 2017 Dec;108:134-145. [doi: 10.1016/j.ijmedinf.2017.10.008] [Medline: 29132619]
95. Esteban B, Tejeda-Lorente, Porcel C, Arroyo M, Herrera-Viedma E. TPLUFIB-WEB: a fuzzy linguistic web system to help in the treatment of low back pain problems. Know Based Syst 2014 Sep;67:429-438. [doi: 10.1016/j.knosys.2014.03.004]
96. Lee C, Wang M, Lan S. Adaptive personalized diet linguistic recommendation mechanism based on type-2 fuzzy sets and genetic fuzzy markup language. IEEE Trans Fuzzy Syst 2015 Oct;23(5):1777-1802. [doi: 10.1109/tfuzz.2014.2379256]
97. Casino F, Patsakis C, Batista E, Borras F, Martinez-Balleste A. Healthy routes in the smart city: a context-aware mobile recommender. IEEE Softw 2017 Nov;34(6):42-47. [doi: 10.1109/ms.2017.4121209]

98. Nurbakova D, Laporte L, Calabretto S, Gensel J. Users Psychological Profiles for Leisure Activity Recommendation: User Study. In: Proceedings of the International Workshop on Recommender Systems for Citizens. 2017 Presented at: RSC'17; August 27, 2017; Como, Italy p. 1-4. [doi: 10.1145/3127325.3127328]
99. Buhl M, Famulare J, Glazier C, Harris J, McDowell A, Waldrip G, et al. Optimizing Multi-channel Health Information Delivery for Behavioral Change. In: IEEE Systems and Information Engineering Design Symposium. 2016 Presented at: SIEDS'16; April 29, 2016; Charlottesville, VA, USA. [doi: 10.1109/sieds.2016.7489285]
100. Chen RC, Ting YH, Chen JK, Lo YW. The Nutrients of Chronic Diet Recommended Based on Domain Ontology and Decision Tree. In: Conference on Technologies and Applications of Artificial Intelligence. 2015 Presented at: TAAI'15; November 20-22, 2015; Tainan, Japan. [doi: 10.1109/taai.2015.7407127]
101. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
102. Massimo D, Elahi M, Ge M, Ricci F. Item Contents Good, User Tags Better: Empirical Evaluation of a Food Recommender System. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. USA: ACM; 2017 Presented at: UMAP'17; July 9-12, 2017; Bratislava, Slovakia p. 373-374. [doi: 10.1145/3079628.3079640]
103. Pilloni P, Piras L, Carta S, Fenu G, Mulas F, Boratto L. Recommendation in Persuasive Ehealth Systems: an Effective Strategy to Spot Users' Losing Motivation to Exercise. In: 2nd International Workshop on Health Recommender Systems. 2017 Presented at: HealthRecSys'17; August 14, 2017; Singapore. [doi: 10.1109/mc.2018.1731060]
104. Herlocker J, Konstan J, Terveen L, Riedl J. Evaluating collaborative filtering recommender systems. ACM Trans Inf Syst 2004 Jan;22(1):5-53. [doi: 10.1145/963770.963772]
105. Pazzani M, Billsus D. Content-based recommendation systems. In: Brusilovsky P, Nejdl W, editors. The Adaptive Web: Methods and Strategies of Web Personalization. Berlin: Springer Berlin Heidelberg; 2007:325-341.
106. Burke R. Knowledge-based recommender systems. Encyclopedia of library and information science 2000;69:1-23.
107. Burke R. Hybrid recommender systems: survey and experiments. User Model User-Adap Inter 2002;12:331-370. [doi: 10.1023/A:1021240730564]
108. Hawker S, Payne S, Kerr C, Hardey M, Powell J. Appraising the evidence: reviewing disparate data systematically. Qual Health Res 2002 Nov;12(9):1284-1299. [doi: 10.1177/1049732302238251] [Medline: 12448672]
109. Adams R, Sadasivam RS, Balakrishnan K, Kinney RL, Houston TK, Marlin BM. PERSPeCT: Collaborative Filtering for Tailored Health Communications. In: Proceedings of the 8th ACM Conference on Recommender Systems. 2014 Presented at: ACM'14; October 6–10, 2014; Foster City, Silicon Valley, CA, USA. [doi: 10.1145/2645710.2645768]
110. MedlinePlus - Health Information from the National Library of Medicine. US National Library of Medicine. URL: https://medlineplus.gov/ [accessed 2019-01-27]
111. Krishnan S, Patel J, Franklin M, Goldberg K. A Methodology for Learning, Analyzing, Mitigating Social Influence Bias in Recommender Systems. In: Proceedings of the 8th ACM Conference on Recommender Systems. 2014 Presented at: ACM'14; October 6–10, 2014; Foster City, Silicon Valley, CA, USA. [doi: 10.1145/2645710.2645740]
112. Wang Z, Huang H, Cui L, Chen J, An J, Duan H, et al. Using natural language processing techniques to provide personalized educational materials for chronic disease patients in china: development and assessment of a knowledge-based health recommender system. JMIR Med Inform 2020 Apr 23;8(4):e17642 [FREE Full text] [doi: 10.2196/17642] [Medline: 32324148]
113. Cacheda F, Carneiro V, Fernández D, Formoso V. Comparison of collaborative filtering algorithms. ACM Trans Web 2011 Feb;5(1):1-33. [doi: 10.1145/1921591.1921593]
114. Adomavicius G, Tuzhilin A. Context-aware recommender systems. In: Ricci F, Rokach L, Shapira B, editors. Recommender Systems Handbook. New York, USA: Springer; 2015.
115. Patients Like Me. URL: https://www.patientslikeme.com/ [accessed 2020-01-29]
116. Health Boards. URL: https://www.healthboards.com/ [accessed 2020-01-29]
117. Knijnenburg BP, Willemsen MC, Gantner Z, Soncu H, Newell C. Explaining the user experience of recommender systems. User Model User-Adap Inter 2012 Mar 10;22(4-5):441-504. [doi: 10.1007/s11257-011-9118-4]
118. Peterson C, Park N, Seligman ME. Orientations to happiness and life satisfaction: the full life versus the empty life. J Happiness Stud 2005 Mar;6(1):25-41. [doi: 10.1007/s10902-004-1278-z]
119. Donnellan MB, Oswald FL, Baird BM, Lucas RE. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. Psychol Assess 2006 Jun;18(2):192-203. [doi: 10.1037/1040-3590.18.2.192] [Medline: 16768595]
120. Przybylski AK, Murayama K, DeHaan CR, Gladwell V. Motivational, emotional, and behavioral correlates of fear of missing out. Comput Hum Behav 2013 Jul;29(4):1841-1848. [doi: 10.1016/j.chb.2013.02.014]
121. Dallery J, Cassidy RN, Raiff BR. Single-case experimental designs to evaluate novel technology-based health interventions. J Med Internet Res 2013 Feb 8;15(2):e22 [FREE Full text] [doi: 10.2196/jmir.2227] [Medline: 23399668]
122. Keyes CL. Mental health in adolescence: is America's youth flourishing? Am J Orthopsychiatry 2006 Jul;76(3):395-402. [doi: 10.1037/0002-9432.76.3.395] [Medline: 16981819]

123. Houston TK, Sadasivam RS, Allison JJ, Ash AS, Ray MN, English TM, et al. Evaluating the QUIT-PRIMO clinical practice ePortal to increase smoker engagement with online cessation interventions: a national hybrid type 2 implementation study. Implement Sci 2015 Nov 2;10(154):1-16 [FREE Full text] [doi: 10.1186/s13012-015-0336-8] [Medline: 26525410]
124. Ricci F, Rokach L, Shapira B. In: Ricci F, Rokach L, Shapira B, Kantor P, editors. Introduction to Recommender Systems Handbook. Boston, UK: Springer; 2011.
125. Uzor S, Baillie L, Htun NN, Smit P. Inclusively designing IDA: effectively communicating falls risk to stakeholders. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2018 Presented at: International Conference on Human-Computer Interaction with Mobile Devices and Services; September 3–6, 2018; Barcelona, Spain p. 281-287. [doi: 10.1145/3229434.3229473]
126. Ferro LS, Walz SP, Greuter S. Towards Personalised, Gamified Systems: an Investigation Into Game Design, Personality and Player Typologies. In: Proceedings of the 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death. 2013 Presented at: ACIE'13; October 1, 2013; Melbourne, VIC, Australia p. 1-6. [doi: 10.1145/2513002.2513024]
127. Swearingen K, Sinha R. Beyond Algorithms: An HCI Perspective on Recommender Systems. In: ACM SIGIR Workshop on Recommender Systems. 2001 Presented at: ACM SIGIR'01; April 9-11, 2001; New Orleans Louisiana USA.
128. Sinha R, Swearingen K. The Role of Transparency in Recommender Systems. In: Extended Abstracts on Human Factors in Computing Systems. 2002 Presented at: CHCS'02; April 20-25, 2002; Minneapolis, Minnesota, USA. [doi: 10.1145/506443.506619]
129. Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. Health Literacy: A Prescription to End Confusion. New York, USA: National Academies Press; 2004.
130. Marx P. Providing Actionable Recommendations. Florida, USA: Josef Eul Verlag GmbH; 2013.
131. Gutiérrez F, Charleer S, De Croon R, Htun NN, Goetschalckx G, Verbert K. Explaining and Exploring Job Recommendations: a User-driven Approach for Interacting With Knowledge-based Job Recommender Systems. In: Proceedings of the 13th ACM Conference on Recommender Systems. 2019 Presented at: ACM'19; September 16–20, 2019; Copenhagen, Denmark. [doi: 10.1145/3298689.3347001]
132. Gutiérrez F, Htun NN, Schlenz F, Kasimati A, Verbert K. A review of visualisations in agricultural decision support systems: an HCI perspective. Comput Electr Agri 2019 Aug;163:104844. [doi: 10.1016/j.compag.2019.05.053]
133. Editorial Policies. Journal of Medical Reseach. URL: https://www.jmir.org/about/editorialPolicies [accessed 2021-06-04]

Abbreviations

DOC: degrees of compromise
HRS: health recommender system
ISO/IEC: International Organization of Standardization/International Electrotechnical Commission
nDCG: normalized discounted cumulative gain
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RQ: research question

Edited by G Eysenbach; submitted 29.01.20; peer-reviewed by A Calero Valdez, J Jiang; comments to author 10.03.20; revised version received 20.05.20; accepted 24.05.21; published 29.06.21.

Please cite as:
De Croon R, Van Houdt L, Htun NN, Štiglic G, Vanden Abeele V, Verbert K
Health Recommender Systems: Systematic Review
J Med Internet Res 2021;23(6):e18035
URL: https://www.jmir.org/2021/6/e18035
doi:10.2196/18035
PMID:34185014

Review

Digital Self-Management Support Tools in the Care Plan of Patients With Cancer: Review of Randomized Controlled Trials

Danielle JM Adriaans[1,2], MSc; Angelique TM Dierick-van Daele[2,3,4], PhD; Marc Johannes Hubertus Maria van Bakel[4], MSc; Grard AP Nieuwenhuijzen[1], MD, PhD; Joep AW Teijink[1], MD, PhD; Fanny FBM Heesakkers[1], MSc; Hanneke WM van Laarhoven[5], MD, PhD

[1]Department of Surgery, Catharina Hospital, Eindhoven, Netherlands
[2]Fontys University of Applied Sciences, Eindhoven, Netherlands
[3]Department of Education and Research, Catharina Hospital, Eindhoven, Netherlands
[4]Fontys School of People and Health Studies, Eindhoven, Netherlands
[5]Department of Medical Oncology, Cancer Center Amsterdam, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Danielle JM Adriaans, MSc
Department of Surgery
Catharina Hospital
Michelangelolaan 2
Eindhoven, 5602 ZA
Netherlands
Phone: 31 624720706
Email: danielle.adriaans@catharinaziekenhuis.nl

Abstract

Background: Digital self-management support tools (DSMSTs)—electronic devices or monitoring systems to monitor or improve health status—have become increasingly important in cancer care.

Objective: The aim of this review is to analyze published randomized clinical trials to assess the effectiveness of DSMSTs on physical and psychosocial symptoms or other supportive care needs in adult patients with cancer.

Methods: Five databases were searched from January 2013 to January 2020. English or Dutch language randomized controlled trials comparing DSMSTs with no intervention, usual care, alternative interventions, or a combination and including patients aged ≥18 years with pathologically proven cancer in the active treatment or survivorship phases were included. The results were summarized qualitatively.

Results: A total of 19 publications describing 3 types of DSMSTs were included. Although the content, duration, and frequency of interventions varied considerably across studies, the commonly used elements included an assessment component, tailored symptom self-management support, an information section, a communication section, and a diary. Significant positive effects were observed on quality of life in 6 (out of 10) studies, on anxiety in 1 (out of 5) study and depression in 2 (out of 8) studies, on symptom distress in 5 (out of 7) studies, on physical activity in 4 (out of 6) studies, on dietary behavior in 1 (out of 4) study, and on fatigue in 2 (out of 5) studies. Moreover, significant negative effects were observed on anxiety in 1 (out of 5) study and depression in 1 (out of 8) study. Most interventions were web-based interventions; 2 studies used mobile apps, and 1 study used a game as a DSMST. The overall quality of the studies was found to be good, with 13 out of 19 studies classified as *high quality*.

Conclusions: This review suggests that DSMSTs have a beneficial effect on the quality of life. For effects on other patient outcomes (eg, anxiety and depression, symptom distress, physical activity, dietary behavior, and fatigue), the evidence is inconsistent and limited or no effect is suggested. Future research should focus on specific tumor types, study different types of interventions separately, and assess the effects of specific interventions at different stages of disease progression.

(J Med Internet Res 2021;23(6):e20861) doi:10.2196/20861

KEYWORDS

web-based intervention; digital self-management support tool; chronic patient groups; review; digital health; ehealth; mhealth; cancer patients; mobile phone

Introduction

Care for patients with cancer extends over a prolonged period, starting with the diagnostic phase, followed by a phase of active treatment and, subsequently, the follow-up phase (in the curative setting), or the supportive care phase (in the palliative setting). Considering the definition of chronic patients by the World Health Organization (patients who require “ongoing management for years or decades covering a wide range of health problems”), in some cases, patients with cancer may be considered as chronic patients [1].

Currently, people with a chronic condition are expected to play a more active role in their health care, which involves symptom management, adherence to treatment regimens, commitment to appropriate lifestyle changes, and the ability to deal with the psychological and physical consequences of their condition [2,3]. Studies related to chronic patients have demonstrated that self-management programs may be associated with reductions in anxiety and unscheduled physician visits and increases in self-efficacy [4-6].

Self-management of chronic disease is challenging for patients, and support from health care professionals is needed. Self-management support is defined as the systematic provision of education and supportive interventions by health care professionals to increase patients’ skills and confidence in managing their health problems, including regular assessment of progress and problems, goal setting, and problem-solving support [6]. Nowadays, it is offered through face-to-face contact and via digital tools.

Digital self-management support tools (DSMSTs) can be any type of electronic device (eg, website and app) or monitoring system (eg, smartwatch) that is applied by physicians in their health care practice or by individuals to monitor or improve their health status. Such tools can be used to stimulate a positive health behavior change, assist individuals to lead a healthier lifestyle, or support the diagnosis and treatment of diseases [7]. DSMSTs provide the means to facilitate communication between health care providers and patients, to transfer information, to improve some clinical outcomes (ie, physical outcome and functional status) among users, and to facilitate patient self-management, thus improving patient empowerment [6,8,9].

Although the population of patients with cancer is growing owing to the aging population and improved cancer care, complaints, needs, and preferences of patients with cancer can vary individually over different subjects and time [10], placing health care budgets under increasing strain. As a result, health authorities are seeking to lessen the burden by using technology to support a move toward self-care and outpatient long-term monitoring. With the rapid development of medical technology in health care, the use of DSMSTs to support patients with cancer will likely become increasingly important and could represent a helpful intervention to enhance psychological well-being (eg, less symptom distress and anxiety) and physical well-being (eg, increasing physical activity [PA]). Despite the projected proliferation of interventions with DSMSTs to manage treatment-related symptoms in patients with cancer, the evidence is lacking and the effectiveness of these tools is still unclear. Previously, researchers reviewed DSMSTs for patients with cancer and found promising results [11-15]. However, these reviews included studies that were primarily focused on cancer survivors [11-13,15] or focused only on a single outcome, that is, patient empowerment or fatigue [12,13,15], or a single digital medium, that is, mobile health [14]. The effects of DSMSTs from a broader perspective, including effects on physical and psychosocial symptoms or other supportive care needs, have not been reviewed before. Therefore, the overall objective of this review is to analyze published randomized clinical trials to assess the effectiveness of DSMSTs on physical and psychosocial symptoms or other supportive care needs in adult patients with cancer.

Methods

Eligibility Criteria for Article Selection

Study Design

Eligible studies were randomized controlled trials (RCTs) in English, performed in adult patients with cancer (≥18 years), published between January 2013 and January 2020, and comparing quantitative physical and/or psychosocial outcomes of DSMSTs with another intervention or usual care. *Patients with cancer* were defined as individuals diagnosed with any type of cancer, irrespective of disease stage, treatment phase, type of treatment, and time since diagnosis. When studies reported on mixed populations, only studies that reported data for patients with cancer separately were included.

Digital Self-Management Support Interventions

Digital self-management support was defined as self-management provided by DSMSTs. To be classified as a self-management support intervention, the intervention should meet criteria 1 and 2:

1. Self-management support targeted at physical or psychosocial symptoms or other supportive care needs of patients: Self-management support is defined as the systematic provision of education and supportive interventions by health care professionals to increase patients’ skills and confidence in managing their health problems, including regular assessment of progress and problems, goal setting, and problem-solving support.
2. A digital tool is used [5].

Programs that were not primarily designed to support or rehabilitate (eg, treatment decision aids and health behavior change interventions) were beyond the scope of this review and were excluded. Programs focusing exclusively on education were only included if the education aimed to support or rehabilitate patients with cancer (eg, group-based, individual-based, structured, and unstructured). Cancer self-management education was defined as an ongoing process of facilitating the knowledge, skills, and confidence necessary to enable effective self-management of the biological, physical, and psychosocial effects of cancer and its treatment [16]. Studies describing interventions without access to the internet or a website were excluded.

Outcomes

Physical parameters related to activity level, dietary behavior, and fatigue and psychosocial parameters (eg, anxiety and depression, quality of life [QOL], and symptom distress) were the outcomes of interest.

Selection Method

To identify potentially relevant studies, CINAHL, Embase, PsycINFO, Cochrane Network, and PubMed databases were searched for eligible RCTs from January 2013 to January 2020. The review began in 2018. Due to the rapid development of medical technology in health care, only studies from the last 5 years were included. During the time of writing this paper, the search was continually updated until January 2020, while maintaining the years 2013 and 2014, given the relevance of the included studies. The search strategy consisted of Medical Subject Headings combined with text words for cancer (Textbox 1) in a Boolean search. A medical information specialist checked the final search syntaxes. DJMA and MJHMVB screened the titles, abstracts, and full texts. Interresearcher reliability was checked using a 20% random sample of abstracts and full texts. Consensus was reached through discussion.

Textbox 1. Medical subject headings and keywords used.

Medical Subject Headings

- self-management, self-management support, self-care, support, supportive care, health services needs and demand, patient education as topic, patient-centered care, health education, action plan, management plan, decision support techniques, continuity of patient care, patient decision making, computer assisted patient decision making, computer assisted decision support system, decision aid*, patient education, patient participation, physician-patient relations, patient information, medical information, decision support, decision tree, decision, decid*, consumer health information, interactive health communication, app, digital health, mobile technology, web based, computer, telemedicine, eHealth, health technology, educational technology, mHealth, mobile phone, smartphone, mobile apps, internet, telecare

Keywords

- cancer, neoplasm*, malignancy, malignancies, tumor

Data Extraction

The following information was extracted from each publication: study characteristics (country of origin, year of publication, aim, and sample size), patient characteristics (age, gender, and type of disease), intervention characteristics (content, duration, and frequency), and outcome measures (instruments used and effects on physical and psychosocial outcomes). The first author independently extracted the data, and another author checked the data extraction for 20% of the studies to determine interrater reliability. Consensus was reached through discussion.

Quality Assessment

The methodological quality of the studies was evaluated, but it did not serve as an eligibility criterion. We used the CONSORT (Consolidated Standards of Reporting Trials) list developed by the CONSORT group to identify the problems arising from inadequate reporting of RCTs [17]. Items were scored using a tick mark. The tick marks indicate "yes" as an answer to the question, resulting in a maximum quality score of 37. For the qualitative synthesis, we counted the overall scores and classified them into 3 quality categories: *high quality* (CONSORT score >25), *moderate quality* (CONSORT score 13-25), and *low quality*. (CONSORT score <13)

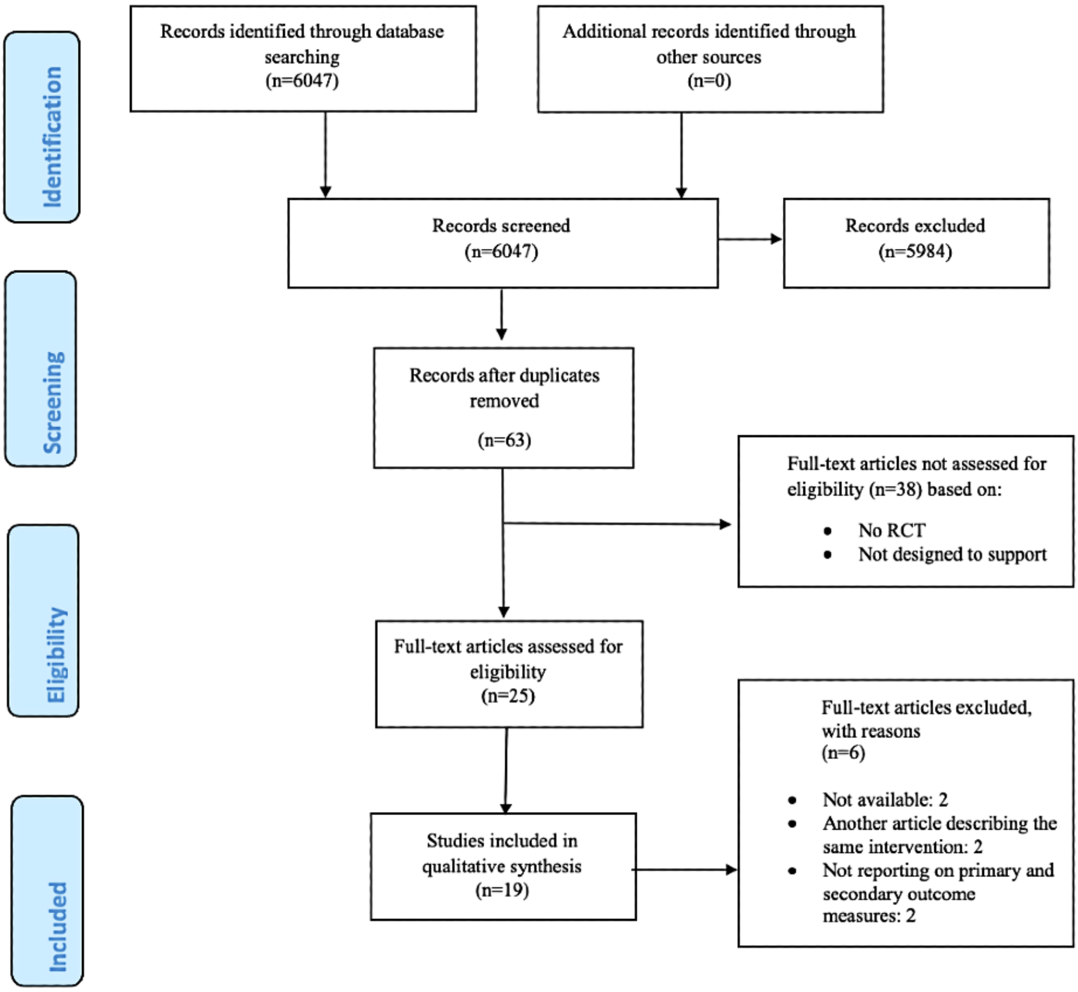
Two reviewers (DJMA and MJHMVB) independently reviewed the papers and independently assessed the methodological quality. In case of disagreement, consensus was reached through discussion.

Results

Selection of Publications

Figure 1 outlines the search process. A total of 6047 references were identified through the search. Screening titles, abstracts, and full texts yielded 19 eligible studies.

Figure 1. Flowchart of included studies. RCT: randomized controlled trial.



Study Characteristics

A total of 19 publications were included (Figure 1); 3 publications were based on the same study, assessing different outcome measures [18-20]. The sample sizes ranged from 39 to 752 patients. All studies had a pre- and posttest design to measure the outcome differences in the groups. One study used 2 experimental groups [21], examining an internet-based patient-provider communication service with or without the additional use of a web-based illness management system, and another study used 2 experimental groups [22]: an unsupervised group that used a mobile app to record data or a supervised group that used the app and reviewed data with a physician. All other studies used a single experimental and control group (eg, control group assigned to a waiting list and received the intervention after the active treatment group, control group receiving written formats, and control group receiving usual care).

Quality of Included Studies

Table 1 presents the methodological quality of the studies. Allocation concealment was described in 9 studies [18-26]. Blinding of participants, personnel, and outcome assessors was adequately described in 8 studies [18,19,22,24,26-29]. Two studies explicitly described a nonblinded approach [21,30]. Out of 37 points, one study achieved 33 (89%) points and had the highest score [29]. A total of 13 studies were of *high quality* and 6 were of *moderate quality* [31].

Table 1. Quality of randomized controlled trials. To make them comparative, overall scores are counted (n, %; maximum score 33; n=37).

Checklist item	[24][a]	[30][b]	[21][c]	[29][d]	[32][e]	[25][f]	[27][g]	[23][h]	[33][i]	[20][j]	[18][k]	[19][l]	[22][m]	[34][n]	[26][o]	[35][p]	[36][q]	[37][r]	[28][s]
1a. Identification as a randomized trial in the title	✓[t]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X[u]	✓	✓	✓	✓	X
1b. Structured summary of trial design, methods, results, and conclusions	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2a. Scientific background and explanation of rationale	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2b. Specific objective or hypotheses	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3a. Description of trial design, including allocation ratio	X	✓	✓	✓	X	✓	X	✓	✓	X	✓	✓	✓	X	✓	✓	✓	✓	X
3b. Important changes to methods after trial commencement	✓	X	✓	✓	X	X	X	X	✓	X	X	X	X	X	X	X	✓	X	X
4a. Eligibility criteria for participants	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4b. Settings and locations where the data were collected	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	X	✓	✓	✓
5. Interventions for each group with sufficient details to allow replication	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X	X	X	X	✓	X	✓	✓	✓
6a. Completely defined prespecified primary and secondary outcome measures	X	✓	✓	✓	✓	✓	X	X	X	✓	✓	✓	✓	X	✓	✓	✓	X	✓
6b. Any changes to trial outcomes after the trial commenced	X	X	X	✓	X	X	X	X	✓	X	X	X	X	X	X	X	✓	X	X
7a. How sample size was determined	✓	✓	X	X	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	X
7b. Explanation of any interim analyses and stopping guidelines	X	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓
8a. Method used to generate the random allocation sequence	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	X	✓
8b. Type of randomization	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	X	X
9. Mechanism used to implement the random allocation sequence	✓	X	✓	X	X	✓	X	✓	X	✓	✓	✓	✓	X	✓	X	X	X	X
10. Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	X	✓	✓	✓	✓	X	X	X	X	X	X	X	X	X	✓	X	X	X	X
11a. Who was blinded after assignment and how	✓	X	X	✓	X	X	✓	X	X	X	✓	✓	✓	X	✓	X	X	X	✓
11b. Description of similarity of interventions	✓	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	X	X	X	✓	✓	✓
12a. Statistical methods used to compare groups for primary and secondary outcomes	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓
12b. Methods for additional analyses	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X
13a. For each group, the numbers of participants who were randomly assigned, who received intended treatment, and who were analyzed for the primary outcome	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Checklist item	[24][a]	[30][b]	[21][c]	[29][d]	[32][e]	[25][f]	[27][g]	[23][h]	[33][i]	[20][j]	[18][k]	[19][l]	[22][m]	[34][n]	[26][o]	[35][p]	[36][q]	[37][r]	[28][s]
13b. For each group, losses and exclusions after randomization, together with reasons	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X
14a. Dates defining the periods of recruitment and follow-up	X	✓	✓	✓	✓	X	X	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	X
14b. Why the trial ended or was stopped	X	X	X	✓	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	X
15. A table showing baseline demographic and clinical characteristics for each group	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓
16. For each group, number of participants included in each analysis and whether the analysis was by original assigned groups	✓	✓	✓	✓	X	✓	✓	X	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓
17a. For each primary and secondary outcome, results for each group and the estimated effect size and its precision were noted	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
17b. For binary outcomes, presentation of both absolute and relative effect sizes is recommended	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
18. Results of any other analyses performed	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X	✓	X	X
19. All important harms or unintended effects in each group	X	X	X	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X	✓	✓	✓
20. Trial limitations	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
21. Generalizability of the trial findings	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X
22. Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
23. Registration number and name of trial registry	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	X	X	✓	✓	✓	X	X
24. Where the full trial protocol can be accessed	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X

Checklist item	[24][a]	[30][b]	[21][c]	[29][d]	[32][e]	[25][f]	[27][g]	[23][h]	[33][i]	[20][j]	[18][k]	[19][l]	[22][m]	[34][n]	[26][o]	[35][p]	[36][q]	[37][r]	[28][s]
25. Sources of funding and other support and role of funders	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

[a] 27 (37).

[b] 30 (81).

[c] 28 (76).

[d] 33 (89).

[e] 25 (68).

[f] 27 (73).

[g] 23 (62).

[h] 28 (76).

[i] 28 (76).

[j] 27 (73).

[k] 29 (78).

[l] 29 (78).

[m] 28 (76).

[n] 15 (41).

[o] 28 (76).

[p] 22 (59).

[q] 29 (78).

[r] 22 (59).

[s] 19 (51).

[t] Reported item.

[u] Unreported item.

Description of Participants

The 19 studies included 5186 patients. Eleven studies included patients in the active treatment phase [21-25,27,28,33-36]. Eight studies included patients who had finished active treatment and were in the curative setting, in the follow-up phase, or in the palliative setting, in the supportive care phase [18-20,26,29,30,32,37]. Nine studies were related to DSMSTs for patients with breast cancer [21,22,24-27,29,32,34]. Six studies were related to patients with cancer in general [18-20,23,33,37]. Two studies included 129 newly diagnosed patients with cancer, of whom 92 were treated for breast cancer [35], and 625 cancer survivors, of which 138 were treated for breast cancer [30]. Two breast cancer studies focused on patients undergoing chemotherapy [22,24]. Of the remaining studies, 1 study focused on 261 patients with primary cancers that had metastasized to the liver and 1 study on 285 patients with nonsmall cell lung cancer [28,36]. The mean number of participants was 273 (range 39-752), of which 70.99% (3682/5186) were female. Some studies included only female participants. The mean age of the subjects was 54.2 years (range 42.35-61.7; Table 2).

Table 2. Characteristics of included studies: population, intervention and comparison descriptions, and study design (N=19).

Reference and country	Population	Stage of care process	Intervention	Technological components	Comparison	Length of intervention	Follow-up
Børøsund et al [21], Norway	167 patients recently diagnosed with breast cancer	Active cancer treatment	Web-based self-management support system and e-messages	Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions[a]; tailored symptom self-management support[b]; information section[c]; communication section[d]; and diary[e]	Care as usual	Minimally 6 months	Baseline, 2 months, 4 months, and 6 months
Ruland et al [25], Norway	325 patients with breast cancer (surgery plus additional treatment) or prostate cancer	Active cancer treatment	Web-based self-management support system	Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions[a]; tailored symptom self-management support[b]; information section[c]; communication section[d]; and diary[e]	Information sheet with relevant internet sites that could be useful to them	1 year	Baseline, 3 months, 6 months, 9 months, and 2 months
Ryhänen et al [27], Finland	300 newly diagnosed patients with breast cancer	Active cancer treatment	Web-based patient education tool	Information section[c]	Usual care: oral and written patient education material	Average 9 months	Just before surgery, 1 day after surgery, when meeting the oncologist for the first time, before and after chemotherapy, before and after radiotherapy, and 1 year after breast cancer diagnosis
Beatty et al [23], Australia	60 patients with cancer	Active cancer treatment	Self-guided, web-based cognitive behavioral therapy	Tailored symptom self-management support[b]	Information-only version of CCO[f]; contained the same 6 information topics as the intervention but no worksheets, activities, relaxation or meditation exercises, or journal	6 weeks	Baseline, immediately postintervention, 3 months postintervention, and 6 months postintervention
Berry et al [33], United States	752 ambulatory adult participants with various cancer diagnoses	Active cancer treatment	Web-based, self-report assessment and educational intervention	Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions[a]; information section[c]; communication section[d]; and diary[e]	Screening for symptom or QOL[g]	From the start of a new therapeutic regimen till 2-4 weeks after treatment ended	Before a new therapeutic regimen, 3-6 weeks after starting treatment, 2 weeks later, and 2-4 weeks after treatment ended
Urech et al [35], Switzerland	129 newly diagnosed patients with cancer (92 treated for breast cancer)	Active cancer treatment	Web-based intervention on stress management	Tailored symptom self-management support[b]	Wait-list control	At least 8 weeks	Baseline, postintervention, and 2-month follow-up

Reference and country	Population	Stage of care process	Intervention	Technological components	Comparison	Length of intervention	Follow-up
Steel et al [28], United States	261 patients diagnosed with hepatocellular, cholangiocarcinoma, gallbladder, neuroendocrine, and pancreatic carcinoma or other primary cancers that have metastasized to the liver	Active cancer treatment	Web-based self-management support system	Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions[a]; tailored symptom self-management support[b]; information section[c]; communication section[d]; and diary[e]	Usual care	6 months	Baseline and 6 months
Gustafson et al [36], United States	285 dyads consisting NSCLC[h] at stage IIIA, IIIB, or IV—patients and a patient-identified primary caregiver	Active cancer treatment	Web-based intervention	Information section[c] and communication section[d]	Training on using the internet and a list of internet sites about lung cancer	25 months or 13 months after patient death, whichever was less	Baseline, 2 months, 4 months, 6 months, and 8 months after the intervention
Egbring et al [22], Switzerland	139 patients with breast cancer undergoing chemotherapy	Active cancer treatment	Mobile app, supervised, and unsupervised	Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions[a]	Usual care	6 weeks	Day 1, day 21, and day 42 during their chemo-therapeutic intervention
Foley et al [34], Ireland	39 patients with breast cancer undergoing surgery	Active cancer treatment	Mobile app	Information section[c]	Not specified	2 weeks	At enrolment, 1 day before surgery, 1 day postsurgery, and 7 days postsurgery
Kim et al [24], Republic of Korea	76 patients with metastatic breast cancer planned to receive chemotherapy	Active cancer treatment	Mobile game	Information section[c]	Usual care+a brochure with side effects of chemotherapy	3 weeks	Baseline and after 3 weeks
Galiano-Castillo et al [26], Spain	81 patients with breast cancer after completing adjuvant therapy	Finished active cancer treatment	Web-based tailored exercise program	Tailored symptom self-management support[b] and communication section[d]	Basic recommendations (written format) for exercise	8 weeks	Baseline, 8 weeks, and 6 months
van den Berg et al [29], the Netherlands	150 female breast cancer survivors 2-4 months before baseline assessment	Finished active cancer treatment	Web-based self-management support system	Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions[a]; tailored symptom self-management support[b]; and information section[c]	Care as usual	4 months	Baseline, 4 months, 6 months, and 10 months

Reference and country	Population	Stage of care process	Intervention	Technological components	Comparison	Length of intervention	Follow-up
Lee et al [32], South Korea	59 patients with breast cancer who had received curative surgery and completed primary cancer treatment within 12 months before the study: diagnose stage 0-III cancers within 2 years before the study	Finished active cancer treatment	Web-based self-management exercise and diet intervention support system	Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions[a]; tailored symptom self-management support[b]; information section[c]; communication section[d]; and diary[e]	Intervention: a 50-page educational booklet on exercise and diet	12 weeks	Baseline and 12 weeks
Van der Hout et al [30], the Netherlands	625 survivors diagnosed with head and neck cancer, colorectal cancer, breast cancer, Hodgkin lymphoma, or non-Hodgkin lymphoma	Finished active cancer treatment	Web-based eHealth app	Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions[a]; tailored symptom self-management support[b]; information section[c]; and communication section[d]	Wait-list control group (access to app after 6 months)	6 months	Baseline, 1 week postintervention, 3 months, and 6 months
Willems et al [20], the Netherlands	462 patients with cancer from 21 different Dutch hospitals	Finished active cancer treatment	Web-based self-management support system	Tailored symptom self-management support[b] and information section[c]	Access to the intervention was postponed until after the 12-month measurement	12 months	Baseline, 3 months, 6 months, and 12 months
Kanera et al [18], the Netherlands	Same intervention as that used by Willems et al [20]	Finished active cancer treatment	Web-based self-management support system	Tailored symptom self-management support[b] and information section[c]	Access to the intervention was postponed until after the 12-month measurement	6 months	Baseline, 3 months, 6 months
Kanera et al [19], the Netherlands	Same intervention as that used by Willems et al [20]	Finished active cancer treatment	Web-based self-management support system	Tailored symptom self-management support[b] and information section[c]	Access to the intervention was postponed until after the 12-month measurement	12 months	Baseline, 3 months, 6 months, and 12 months

Reference and country	Population	Stage of care process	Intervention	Technological components	Comparison	Length of intervention	Follow-up
Bantum et al [37], United States	352 cancer survivors	Finished active cancer treatment	Web-based self-management support system	Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions[a]; tailored symptom self-management support[b]; information section[c]; communication section[d]; and diary[e]	Delayed-treatment control condition	6 months	Baseline and 6 months

[a] An assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions, currently and over time.

[b] Tailored symptom self-management support to self-manage symptoms and problems the patient experiences.

[c] An information section, which included information about various aspects of cancer such as exercise, nutrition, coping, and symptom management and also provided access to other reliable and relevant web sources.

[d] Communication section, with fellow patients or with health care providers, using discussion centers, an SMS text messaging function, or email as a communication tool.

[e] Diary, where patients could keep personal notes.

[f] CCO: Cancer Coping Online.

[g] QOL: quality of life.

[h] NSCLC: nonsmall cell lung carcinoma.

Intervention Characteristics

The intervention characteristics for both the intervention and control groups are described in Table 2. The degree of detail provided about the interventions varied greatly across studies. There was a large variation in the duration, frequency, and content of the interventions. The mean duration of the intervention was 39.5 weeks (range 2 weeks to 25 months). A total of 37% (7/19) interventions focused only on the psychological well-being of patients [27,29,30,33-36], 1 focused only on physical health [32], and 11 focused on both [18-26,28,37].

The technological component was mainly a web-based approach (16/19, 84%) [18-21,23,25-30,32,33,35-37]; in 2 studies, a mobile app was used [22,34], and 1 study used a mobile game as a DSMST [24]. Of the 16 studies that used a web-based approach, 1 study sent email reminders in an attempt to maintain or improve adherence [29].

Although the content of the interventions differed, 5 key components of DSMSTs were identified (Table 2). A total of 5 of the 16 web-based approach studies used all 5 key components in their DSMSTs [21,25,28,29,32,37] to increase self-management (support): (1) An assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions, currently and over time (eg, improving diet, increasing exercise, and stress management via relaxation therapy); (2) Tailored symptom self-management support to self-manage symptoms and problems the patient experiences (eg, in the study by Børøsund et al [21], patients could choose symptoms and problems they were experiencing from a predefined list, rate the burden of these symptoms and problems, and indicate where they needed help); (3) The information section, which included information about various aspects of cancer such as exercise, nutrition, coping, and symptom management and also provided access to other reliable and relevant web sources; (4) Communication section, with fellow patients or with health care providers, using discussion centers, an SMS text messaging function, or email as a communication tool. Communication with fellow patients was often used for social networking, providing feedback, and encouraging each other, whereas communication with health care providers was often used for difficult questions and support; (5) Diary, where patients could keep personal notes. The 2 studies that provided a mobile app only offered 1 of the 5 key components. One offered tailored information and the other offered an assessment component (Table 2) [22,34]. One study, using a mobile game, offered patient education as a key component to increase the self-management of patients with breast cancer [24].

Outcomes of Included Studies

The measurement instruments used and the corresponding outcomes of the studies are presented in Table 3. Psychosocial outcome measures, such as QOL, anxiety and depression, and symptom distress, were the most commonly used outcome measures, mostly using validated (eg, The European Organisation for Research and Treatment of Cancer QOL Questionnaire Core 30, Functional Assessment of Cancer Therapy-Fatigue, Hospital Anxiety and Depression Scale [HADS]) questionnaires.

Table 3. Intervention outcomes.

Reference and country	Outcomes and measurement instruments	Results					
		QOL[a]	Anxiety and depression	Distress (symptom)	Fatigue	PA[b]	Dietary behavior
Børøsund et al [21], Norway	Anxiety, depression, (HADS[c]), and symptom distress (MSAS[d])	—[e]	• Anxiety: intervention+communication service<control (P=.03) • Intervention+communication service versus communication service: NS[f] • Depression: intervention+communication service<control group (P=.03) • Intervention+communication service versus communication service: NS • Depression: communication service<control (P=.03)	• Intervention+communication service<control (P=.001) • Intervention+communication service versus communication service: NS	—	—	—
Ruland et al [25], Norway	Symptom distress (MSAS), depression (Center for Epidemiological Studies-Depression Scale), self-efficacy, and social support	• Intervention=control (P=.18)	• Depression: intervention=control (P=.16)	• Intervention<control (P=.04; only on global symptom distress index)	—	—	—
Ryhänen et al [27], Finland	QOL (QOL-CS[g]), anxiety (STAI[h]), and side effects	• Intervention=control (P=.82)	• Anxiety: intervention=control (P=.64)	—	—	—	—

Reference and country	Outcomes and measurement instruments	Results					
		QOL[a]	Anxiety and depression	Distress (symptom)	Fatigue	PA[b]	Dietary behavior
Beatty et al [23], Australia	Distress (PSS-SR[i]; DASS[j]), HRQOL[k] (EORTC-QLQ-C30[l]), and coping (mini-MAC[m])	• Intervention>control at 3-month follow-up • Intervention>control at 6-month follow-up for global QOL (*d*=−0.43) • Trend toward a significant group×time interaction for global QOL	—	• Intervention=control at 3-month follow-up	—	• Intervention>control • At 3-month follow-up (*d*=−0.52; *P*=.02)	—
Berry et al [33], United States	Symptom distress (SDS-15[n] score)	—	—	• The SDS-15 score was reduced by an estimated 1.53 points (*P*=.01) in the intervention group users compared with the matched control group.	—	—	—
Urech et al [35], Switzerland	QOL (FACIT-F[o]), anxiety or depression (HADS), and distress (distress thermometer)	• Intervention>control (*P*=.007)	• Intervention=control (*P*=.15)	• Intervention<control (*P*=.03) • Immediately after the intervention • After 2 months, intervention=control	—	—	—
Steel et al [28], United States		• Intervention: QOL increased (Cohen *d*=0.99)	• Intervention: depression decreased (*d*=0.71)	—	• Intervention: NS	—	—

Reference and country	Outcomes and measurement instruments	Results					
		QOL[a]	Anxiety and depression	Distress (symptom)	Fatigue	PA[b]	Dietary behavior
	Depression (Center for Epidemiological Studies-Depression), pain (BPI[p]), fatigue (FACT[q] instrument), HRQOL (FACT-G[r]), and caregiver stress and depression (CQOLC[s] and Center for Epidemiological Studies-Depression scale)						
Gustafson et al [36], United States	Symptom distress (ESAS[t])	—	—	- Intervention<control - Significant differences at 4 months (*P*=.03; *d*=0.42) and 6 months (*P*=.004; *d*=0.61) - Similar but marginally significant effects were observed at 2 months (*P*=.05; *d*=0.39) and 8 months (*P*=.06; *P*=.43)	—	—	—
Egbring et al [22], Switzerland	Daily functional activity (ECOG[u])	—	—	—	—		—

Reference and country	Outcomes and measurement instruments	Results					
		QOL[a]	Anxiety and depression	Distress (symptom)	Fatigue	PA[b]	Dietary behavior
						• Decreased; All groups from first to second visit • Increased; Intervention: supervised from second to third visit • Decreased; Intervention: unsupervised and control • Intervention: supervised from first (median 90.85, IQR 30.67) to third visit (median 84.76, IQR 18.29; *P*=.72)	
Foley et al [34], Ireland	Anxiety and depression (HADS)	—	• Control<intervention 7 days postoperative (*P*=.03, anxiety; *P*=.02; depression)	—	—	—	—
Kim et al [24], Republic of Korea	QOL (WHO QOL-BREF[v] Scale), anxiety (Spielberger State-Trait anxiety scale), and depression (BDI[w])	• Intervention>control (*P*=.01)	• Anxiety: intervention=control (*P*=.21) • Depression: intervention=control (*P*=.99)	—	—	—	—
Galiano-Castillo et al [26], Spain	QOL (EORTC-QLQ-C30) and fatigue (R-PFS[x])		—	—	• Fatigue • Intervention<control (*P*<.001)	—	—

Reference and country	Outcomes and measurement instruments	Results					
		QOL[a]	Anxiety and depression	Distress (symptom)	Fatigue	PA[b]	Dietary behavior
		Intervention>control for global health status (*P*=.001), physical functioning (*P*=.001), role functioning (*P*=.003), and cognitive functioning (*P*=.007), and arm symptoms (*P*=.003),					
Berg et al [29], the Netherlands	Distress (SCL-90[y])	—	—	• Intervention<control (*P*=.02)	—	—	—
Lee et al [32], South Korea	HRQOL (EORTC-QLQ-C30), exercise and intake of Fruit and vegetables, diet quality (DQI[z]), stage of change for exercise, and fatigue (BFI[aa])	• Intervention>control (*P*=.02)	—	—	• Intervention<control (*P*=.03)	• Moderate-intensity aerobic exercise: intervention>control (*P*<.001) • Physical functioning: intervention>control (*P*=.02) • Stage of change for exercise: intervention>control (*P*<.001)	• Overall diet quality: intervention>control (*P*=.001) • Appetite loss: intervention>control (*P*=.03) • Fruit and vegetables consumption: intervention>control (*P*=.03)
Van der Hout et al [30], the Netherlands	HRQOL (EORTC-QLQ including tumor-specific symptoms within the tumor groups)	• Intervention>control over time (*P*=.05)	—	—	—	—	—
Willems et al [20], the Netherlands	Emotional and social functioning (EORTC-QLQ-C30), depression (HADS), and fatigue (CIS[ab])	• Emotional and social functioning • 6 months: intervention ↑ social functioning in men (*d*=0.34) • 12 months: intervention=control	• 6 months: intervention ↓ for participants who received chemotherapy (*d*=0.36) • 12 months: intervention=control	—	• 6 months: intervention: a decrease for participants ≤56 years (*d*=0.44) • 12 months: intervention=control	—	—
		—	—	—	—		

Reference and country	Outcomes and measurement instruments	Results					
		QOL[a]	Anxiety and depression	Distress (symptom)	Fatigue	PA[b]	Dietary behavior
Kanera et al [18], the Netherlands	PA (SQUASH[ac]) and dietary behavior (Dutch Standard Questionnaire on Food Consumption)					- Moderate PA intervention>control (*P*<.001) - After multiple testing, significance expired	- Intervention>control (*P*=.02) - After multiple testing, significance expired
Kanera et al [19], the Netherlands	Moderate PA (SQUASH) and vegetable consumption (Dutch Standard Questionnaire on Food Consumption)	—	—	—	—	- Intervention>control (*P*=.01) - Age only significant moderator (*P*=.01)	- Vegetable consumption: intervention=control (*P*=.12)
Bantum et al [37], United States	Fatigue (BFI), exercise (Godin Exercise Questionnaire), fruit and vegetable intake (Block Food Frequency Questionnaire), and depression (PHQ[ad]-8)	—	- Depression: intervention=control (*P*=.69)	—	- Intervention=control (*P*=.56)	- Intervention>control (*P*=.01), increase of strenuous exercise (32-51 min per week compared with a steady 29 min per week) - Intervention>control (*P*=.01), increase of stretching (31 min at baseline to 46 min per week in the intervention group compared with 26 min at baseline to 25 min after 6 months in the control group)	- Fruit and vegetable intake: intervention=control (*P*=.24)

---

[a] QOL: quality of life.

[b] PA: physical activity.

[c] HADS: Hospital Anxiety and Depression Scale.

[d] MSAS: Memorial Symptom Assessment Scale.

[e] Not available.

[f] NS: nonsignificant.

[g] QOL-CS: Quality of Life Cancer Survivor.

[h] STAI: State-Trait Anxiety Inventory.

[i] PSS-SR: Posttraumatic Stress Scale-Self-Report.

[j] DASS: Depression Anxiety Stress Scale.

[k] HRQOL: health-related quality of life.

[l] EORTC-QLQ-C30: The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30.

[m] mini-MAC: mini Mental Adjustment to Cancer Scale.

[n] SDS-15: 15-item Symptom Distress Scale.

[o] FACIT-F: Functional Assessment of Cancer Therapy-Fatigue.

[p] BPI: Brief Pain Inventory.

[q] FACT: Functional Assessment of Cancer Therapy.

[r] FACT-G: Functional Assessment of Cancer Therapy-General.

[s] CQOLC: Caregiver Quality of Life Index-Cancer Scale.

[t] ESAS: Edmonton Symptom Assessment Scale.

[u] ECOG: Everyday Cognition.

[v] WHO QOL BREF scale: World Health Organization Quality of Life-BREF Scale.

[w] BDI: Beck Depression Inventory.

[x] R-PFS: Piper Fatigue Scale-Revised.

[y] SCL-90: Symptom Checklist-90.

[z] DQI: Diet Quality Index.

[aa] BFI: Brief Fatigue Inventory.

[ab] CIS: Checklist Individual Strength.

[ac] SQUASH: Short Questionnaire to Assess Health-Enhancing Physical Activity.

[ad] PHQ: Patient Health Questionnaire.

A total of 10 studies reported QOL [20,23-28,30,32,35], whereas 6 of them reported positive outcomes [24,26,28,30,32,35]. A few studies (4/9, 44%) used The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 to measure QOL. Four studies observed that the overall health-related QOL improved to a significantly larger degree compared with the control group [24,28,30,35]. Two of these studies [24,30] were of *high quality*, including 76 and 625 participants, respectively. One study, including 81 patients who finished active treatment, with 8 weeks of access to an internet-based tailored exercise program, found that health-related QOL improved to a significantly larger degree compared with the control group on the subdomains global health status, physical functioning, role functioning, cognitive functioning, and arm symptoms [26]. One study, which reported on 59 patients with breast cancer who finished active treatment, only found a statistically significant difference in the subdomain physical functioning [32]. Four studies [20,23,25,27] found no statistically significant differences in the overall QOL or subdomains of QOL. Three of these studies included >300 patients each, of which 2 studies were of *high quality*.

Anxiety was reported in 5 studies, whereas depression was reported in 8 studies. Four studies [21,24,34,35] reported anxiety and depression. Four studies used the HADS [20,21,34,35] to report anxiety and depression. Willems et al [20] used only the depression subscales of the HADS. Two other studies also reported depression using different questionnaires [28,37]. One study only reported the outcomes of anxiety [27]. Two studies reported significant differences in anxiety levels [21,34], and 3 studies reported significant differences in depression levels [21,28,34]. One of these studies, including 167 patients with breast cancer in active treatment with access to a web-based self-management support system with e-messages, observed significantly lower anxiety (P=.03) and depression (P=.03) levels in the intervention group than in the usual care group [21]. This study was classified as *high quality*. Another study, classified as *moderate quality* and including 261 active treatment patients, reported a reduction in depression (Cohen d=0.71) for the intervention group when compared with the usual care group [28]. In contrast, another *moderate quality* study, including 39 patients with breast cancer with access to a mobile app for 2 weeks, reported significantly lower anxiety (P=.02) and depression (P=.03) levels in the control group than in the intervention group [34]. Of the studies that found no significant differences in anxiety, 2 studies were of *moderate quality*, whereas 1 was classified as *high quality*, including 76 patients who completed active cancer treatment. Of the studies that found no significant differences in depression, 2 studies were of *moderate quality*, whereas 3 [20,24,25] were classified as *high quality*. One of these studies included 76 patients who completed active cancer treatment and 2 of these studies included 325 patients, during their active cancer treatment.

A total of 7 studies reported symptom distress [21,23,25,29,33,35,36], of which 6 were classified as *high-quality* studies. One study [35] was classified as *moderate quality*. All studies used a web-based approach. One study (including 167 active treatment patients), using 2 experimental groups [26] and examining an internet-based patient-provider communication service with and without the additional use of a web-based illness management system, found significantly lower symptom distress in the web-choice intervention group than in the control group, but no statistically significant differences were observed between the 2 intervention groups [21]. In addition, 2 other studies, including 150 and 752 patients, found significantly less distress in the intervention group [29,33]. Another study, including 325 patients, found significant group differences in symptom distress [25]. One study included 285 active treatment patients and their primary caregivers [36]. These caregivers reported lower patient physical symptom distress in the intervention group than in the control group. One study found that distress was significantly lower immediately after the intervention in the intervention group than in the control group. However, distress did not change significantly from immediately after the intervention to the follow-up 2 months later [35]. One study, including 60 patients, found no statistically significant group×time interactions [23].

Five studies reported fatigue [20,26,28,32,37], of which 2 reported a decrease in fatigue after 8 and 12 weeks of a web-based intervention [26,32]. One of these studies, including 81 patients, was classified as *high quality*. Three others found no significant changes after an intervention of 6 months [20,28,37], of which 1 was classified as *high quality*, including 462 patients.

Six studies reported results on PA [18,19,22,23,32,37]. Four studies observed significant effects; 2 studies were classified as *high quality* [19,23], whereas 2 were classified as *moderate quality* [32,37]. In 1 study, 139 participants were randomly assigned to an unsupervised group (intervention), a supervised group (intervention), or a control group [22]. The intervention groups showed no significant differences from the first to the third visit. On the other hand, another study including 352 patients who finished active treatment with 6 months of access to a web-based self-management support system, showed an increase in strenuous exercise in the intervention group compared with the control group [37]. Another study found that moderate-intensity aerobic exercise for at least 150 minutes per week significantly increased in the intervention group compared with controls [32]. In the study by Beatty et al [23], 60 participants received either the 6-week intervention Cancer Coping Online (n=30) or the 6-week web-based attention control (n=30). The Cancer Coping Online participants had significantly higher physical functioning than the controls at 3 months of follow-up ($d=-0.52$; $P=.02$). The study that found no significant effect [18] was similar to a study that found a significant effect, with the only difference in follow-up time (6 months vs 12 months) [19].

Four studies reported dietary behaviors [18,19,32,37]. One study, including 59 patients with 12 weeks of access to a web-based self-management exercise and diet intervention support system, showed a greater improvement in overall diet quality in the intervention group ($P=.001$) [32]. Another study, including 352 patients who finished active treatment with 6 months of access to a web-based self-management support system, reported no significant changes in fruit and vegetable intake [37]. Both studies were classified as *moderate quality*. Two *high-quality* studies of 462 patients, who completed active treatment, found no significant changes in dietary behavior and vegetable intake in particular [18,19].

Discussion

Principal Findings

In this paper, we have systematically reviewed published RCTs to assess the effectiveness of existing interventions with DSMSTs on physical and psychosocial symptoms or other supportive care needs in adult patients with cancer. A total of 19 publications covering 17 unique studies were included in this review. The RCTs varied in terms of content, duration, and frequency. Nevertheless, we identified 5 elements that were common for the majority of the interventions: an assessment component, tailored symptom self-management support, an information section, a communication section, and a diary. Significant positive effects were observed on QOL in 6 studies, on anxiety in 1 study and depression in 2 studies, on symptom distress in 5 studies, on PA in 4 studies, on dietary behavior in 1 study, and on fatigue in 2 studies. In addition, significant negative effects were observed on anxiety and depression in 1 study. Other studies reported no significant effects on these outcomes (4 studies on QOL, 3 studies on anxiety, 5 studies on depression, 2 studies on symptom distress, 2 studies on PA, 3 studies on dietary behavior, and 3 studies on fatigue). Most interventions were web-based; 2 studies used mobile apps, and 1 study used a game as a DSMST. No relationship was found between the effectiveness of the studies and different components of the DSMSTs for patients with cancer, the quality of the study, and the impact on physical and psychosocial symptoms or other supportive care needs in adult patients with cancer.

Concerning the reported effects of DSMSTs on psychosocial and physical symptoms or other supportive care needs, comparable reviews also showed positive effects of DSMSTs. A systematic review [38] included 16 studies that examined internet-based support programs in patients with cancer. That review showed that internet-based support programs are effective in improving psychosocial and physical symptoms in patients with cancer [38]. Another systematic review included 17 studies that examined web-based mental health interventions in patients with chronic gastrointestinal conditions. That review showed that these interventions resulted in fewer somatic symptoms and improved QOL [39]. Moreover, another review suggested that DSMSTs could be useful for individuals during and after cancer treatment, especially in terms of information, follow-up planning, and management of side effects [14]. However, significant negative effects of DSMSTs were also observed in the reviewed studies. Several studies have reported no significant effects of DSMSTs on specific psychosocial and physical symptoms or other supportive care needs. Other factors may also have played a role in the large variation in observed outcomes. These include the different measurement instruments used within and between studies, different sample sizes, and different periods between the start of the intervention and the postintervention measurement. Future studies should preferably use uniform outcome measures and time intervals for the assessment of outcomes.

For some patients, having more knowledge about their condition might reduce their anxiety as a result of the development of realistic expectations of the future and preparedness for treatment-related side effects. On the contrary, information might also increase patients' anxiety by drawing attention to their condition, unknown symptoms, or risks of treatment. In our review, one study [21] reported significantly lower anxiety and depression levels in the intervention group than in the control group, whereas another study [34] reported significantly higher anxiety and depression levels in the intervention group than in the control group. In the first study, patients had access to the intervention for 1 year and could use the system as much as they liked. In the second study, patients received a tablet computer 1 week before surgery and had to return the tablet 1 week postoperatively. However, the heterogeneity in content, frequency, and duration of the interventions included in our review precludes a definitive answer to the question on the effect of digital self-management support on anxiety and depression. An earlier review on web-based interventions for type 2 diabetes indicated that interventions of longer duration (>12 weeks) resulted in better outcomes. This may also be the case in patients with cancer [40]. However, further studies are needed to confirm this.

We considered the mode of delivery (how the intervention was delivered to the recipients) in the included studies and identified that the technological component was mainly a web-based approach. Over the past 10 years, the number of publications reporting on the use of DSMSTs in health care has increased. At the beginning of this period, studies focused on telehealth, whereas in the past 5 years, the majority of studies reported on the use of eHealth and mobile health in DSMSTs [7]. Mobile devices have emerged as an important tool for enhancing communication between patients and health care providers, for patient engagement in their health, for disease prevention, and for interventions that change health behavior [13,41]. Surprisingly, we found only 2 studies using mobile health apps, which is in contrast with the rapidly growing market of mobile health apps in general. This may imply that the introduction of new mobile health tools is much faster than its scientific appraisals [7]. To better understand why one device in a care program is more effective than others, adequately conducted studies that moderate the possible effects are needed.

We identified 5 elements that were common for the majority of the interventions: an assessment component, tailored symptom self-management support, an information section, a communication section, and a diary. These elements were used as single-component interventions or multicomponent interventions using different combinations of elements. Of the 16 web-based approach studies, 5 studies used all 5 elements in their DSMSTs [21,25,28,29,32,37] to increase self-management (support). Given the design of the studies, it is difficult to determine whether multicomponent DSMSTs are more effective than single-component DSMSTs and, in the case of a multicomponent DSMST, which particular component contributes most to a certain effect [13]. An earlier review of cancer survivors supported the benefit of an educational element, that is, cancer survivors who received sufficient information reported a better QOL [42]. Most reviews targeting cancer [11-15] highlighted that DSMSTs are mostly multicomponent and that there is a great deal of heterogeneity in the protocols and outcomes measured in cancer - related DSMST studies. Future studies should be more structured to determine the role of individual intervention elements and should take the duration and frequency of interventions into account [13]. To further demonstrate the effects in patients, researchers should analyze and compare single-component and multicomponent DSMSTs separately.

Individuals have different preferences regarding information seeking, health care participation, and embracement of DSMSTs. Preferences of women and men might differ regarding health information seeking and support [43]. In addition, age might be of influence, as it might be more difficult to work with new technologies for the older adults [7]. Other factors that could influence the use of new technology are the educational level or the skills needed for using electronic devices [7]. The studies described in this review included patients with different disease or treatment phases. The effects of DSMSTs on physical and psychosocial symptoms or other supportive care needs might differ depending on the patients' need for information and support, which may vary during the phases. Patients with cancer in the curative phase, for example, may need more information on how to cope with the late effects of surgery or chemotherapy, whereas patients in the palliative phase may want information about the self-management of pain and psychological distress [44]. Future reviews should focus more on comparing the effects of DSMSTs in different groups of patients, distinguished by treatment stage (curative or palliative) and tumor types.

Limitations

Although our review was systematic and we took care to assure its quality, there are limitations to our study. One limitation of this review is that the studies included in the review were conducted predominantly with patients with breast cancer. In addition, several studies included in this review enrolled patients with mixed cancer populations. In some cases, the reported effects and evidence found in the included studies may apply more to one type of patients with cancer than to patients with other tumor types. The preferences and needs of patients with a specific tumor type may differ. Therefore, future studies should focus on specific tumor types. Another limitation is that although the average rating for methodology was good (13 studies were of high quality), the trials included in our review had several potential sources of bias and error. In particular, insufficient information regarding allocation concealment and the lack of blinding participants and personnel as well as outcome assessors might have biased the results. A third limitation is that we summarized different types of self-management support interventions for different types of patient groups and compared their benefits for patient self-management. This heterogeneity hampers firm conclusions regarding the effects on the studied outcomes. In addition, some studies comprised small sample sizes [24,26,32,34]. The absence of significant effects might be caused by a lack of power instead of the true ineffectiveness of the intervention. Due to the large variety of outcome measures, study characteristics, and components of DSMSTs, neither meta-analysis nor a comprehensive description of effect sizes was possible.

Conclusions

In conclusion, this review suggests that DSMSTs have a beneficial effect on the QOL. For effects on other patient outcomes (eg, anxiety and depression, symptom distress, PA, dietary behavior, and fatigue), the evidence is inconsistent and limited or no effect is suggested. A total of 5 elements that were common for the majority of the interventions included an assessment component, tailored symptom self-management support, an information section, a communication section, and a diary. We identified several lacunas in the available body of evidence regarding the effects of DSMSTs on patients with specific tumor types, patients with cancer in a specific treatment or disease stage, the design of technology, and especially the design of technology tailored to the patients' needs. Future research should focus on specific tumor types, study different types of interventions separately, and assess the effects of specific interventions at different stages of disease progression.

Authors' Contributions

DJMA was the lead reviewer and drafted the manuscript. MJHMVB was the secondary reviewer for full-text screening, for data extraction, and for quality control. DJMA completed the data analysis and ATMDVD reviewed for accuracy. All authors provided feedback on the draft manuscript. All authors reviewed the final manuscript before submission.

Conflicts of Interest

HWMvL has a consultant or advisory role with BMS, Lilly, MSD, Nordic Pharma, Servier, and receives research funding/medication from Bayer, BMS, Celgene, Janssen, Lilly, Merck, Nordic Pharma, Philips, Roche, Servier. The other authors have no conflict to declare.

References

1. World Health Organization. Global Status Report on Noncommunicable Diseases 2014. Geneva: World Health Organization; 2014:1-298.
2. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. J Am Med Assoc 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
3. Van Hecke A, Heinen M, Fernández-Ortega P, Graue M, Hendriks JM, Høy B, et al. Systematic literature review on effectiveness of self-management support interventions in patients with chronic conditions and low socio-economic status. J Adv Nurs 2017 Apr;73(4):775-793. [doi: 10.1111/jan.13159] [Medline: 27653960]
4. Gao WJ, Yuan CR. Self-management programme for cancer patients: a literature review. Int Nurs Rev 2011 Sep;58(3):288-295. [doi: 10.1111/j.1466-7657.2011.00907.x] [Medline: 21848773]
5. Howell D, Harth T, Brown J, Bennett C, Boyko S. Self-management education interventions for patients with cancer: a systematic review. Support Care Cancer 2017 Apr;25(4):1323-1355. [doi: 10.1007/s00520-016-3500-z] [Medline: 28058570]
6. Institute Of Medicine, Board on Health Care Services, Committee on the Crossing the Quality Chasm: Next Steps Toward a New Health Care System. In: Adams K, Greiner AC, Corrigan JM, editors. The 1st Annual Crossing the Quality Chasm Summit: A Focus on Communities: Report of a Summit. Washington, DC: The National Academies Press; 2004:1-176.
7. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5:290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
8. Alpay LL, Henkemans OB, Otten W, Rövekamp TA, Dumay AC. E-health applications and services for patient empowerment: directions for best practices in The Netherlands. Telemed J E Health 2010 Sep;16(7):787-791. [doi: 10.1089/tmj.2009.0156] [Medline: 20815745]
9. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
10. Kanera IM, Willems RA, Bolman CA, Mesters I, Zambon V, Gijsen BC, et al. Use and appreciation of a tailored self-management eHealth intervention for early cancer survivors: process evaluation of a randomized controlled trial. J Med Internet Res 2016;18(8):e229 [FREE Full text] [doi: 10.2196/jmir.5975] [Medline: 27554525]
11. Goode AD, Lawler SP, Brakenridge CL, Reeves MM, Eakin EG. Telephone, print, and web-based interventions for physical activity, diet, and weight control among cancer survivors: a systematic review. J Cancer Surviv 2015 Mar 11:660-682. [doi: 10.1007/s11764-015-0442-2] [Medline: 25757733]
12. Hong Y, Peña-Purcell NC, Ory MG. Outcomes of online support and resources for cancer survivors: a systematic literature review. Patient Educ Couns 2012 Mar;86(3):288-296. [doi: 10.1016/j.pec.2011.06.014] [Medline: 21798685]
13. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013;15(2):e37 [FREE Full text] [doi: 10.2196/jmir.2281] [Medline: 23425685]
14. Odeh B, Kayyali R, Nabhani-Gebara S, Philip N. Optimizing cancer care through mobile health. Support Care Cancer 2015 Jul;23(7):2183-2188. [doi: 10.1007/s00520-015-2627-7] [Medline: 25649121]
15. Seiler A, Klaas V, Tröster G, Fagundes CP. eHealth and mHealth interventions in the treatment of fatigued cancer survivors: a systematic review and meta-analysis. Psychooncology 2017 Sep;26(9):1239-1253. [doi: 10.1002/pon.4489] [Medline: 28665554]
16. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Fischl AH, et al. Diabetes self-management education and support in Type 2 diabetes: a joint position statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics. J Acad Nutr Diet 2015 Aug;115(8):1323-1334. [doi: 10.1016/j.jand.2015.05.012] [Medline: 26054423]

17. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
18. Kanera IM, Bolman CA, Willems RA, Mesters I, Lechner L. Lifestyle-related effects of the web-based Kanker Nazorg Wijzer (Cancer Aftercare Guide) intervention for cancer survivors: a randomized controlled trial. J Cancer Surviv 2016 Oct 10;10(5):883-897 [FREE Full text] [doi: 10.1007/s11764-016-0535-6] [Medline: 26984534]
19. Kanera IM, Willems RA, Bolman CA, Mesters I, Verboon P, Lechner L. Long-term effects of a web-based cancer aftercare intervention on moderate physical activity and vegetable consumption among early cancer survivors: a randomized controlled trial. Int J Behav Nutr Phys Act 2017 Feb 10;14(1):19 [FREE Full text] [doi: 10.1186/s12966-017-0474-2] [Medline: 28187725]
20. Willems RA, Mesters I, Lechner L, Kanera IM, Bolman CA. Long-term effectiveness and moderators of a web-based tailored intervention for cancer survivors on social and emotional functioning, depression, and fatigue: randomized controlled trial. J Cancer Surviv 2017 Dec;11(6):691-703 [FREE Full text] [doi: 10.1007/s11764-017-0625-0] [Medline: 28698999]
21. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. J Med Internet Res 2014;16(12):e295 [FREE Full text] [doi: 10.2196/jmir.3348] [Medline: 25525672]
22. Egbring M, Far E, Roos M, Dietrich M, Brauchbar M, Kullak-Ublick GA, et al. A mobile app to stabilize daily functional activity of breast cancer patients in collaboration with the physician: a randomized controlled clinical trial. J Med Internet Res 2016 Sep 06;18(9):e238 [FREE Full text] [doi: 10.2196/jmir.6414] [Medline: 27601354]
23. Beatty L, Koczwara B, Wade T. Evaluating the efficacy of a self-guided web-based CBT intervention for reducing cancer-distress: a randomised controlled trial. Support Care Cancer 2016 Mar;24(3):1043-1051. [doi: 10.1007/s00520-015-2867-6] [Medline: 26248651]
24. Kim HJ, Kim SM, Shin H, Jang J, Kim YI, Han DH. A mobile game for patients with breast cancer for chemotherapy self-management and quality-of-life improvement: randomized controlled trial. J Med Internet Res 2018 Oct 29;20(10):e273 [FREE Full text] [doi: 10.2196/jmir.9559] [Medline: 30578205]
25. Ruland CM, Andersen T, Jeneson A, Moore S, Grimsbø GH, Børøsund E, et al. Effects of an internet support system to assist cancer patients in reducing symptom distress: a randomized controlled trial. Cancer Nurs 2013;36(1):6-17. [doi: 10.1097/NCC.0b013e31824d90d4] [Medline: 22495503]
26. Galiano-Castillo N, Cantarero-Villanueva I, Fernández-Lao C, Ariza-García A, Díaz-Rodríguez L, Del-Moral-Ávila R, et al. Telehealth system: a randomized controlled trial evaluating the impact of an internet-based exercise intervention on quality of life, pain, muscle strength, and fatigue in breast cancer survivors. Cancer 2016 Oct 15;122(20):3166-3174 [FREE Full text] [doi: 10.1002/cncr.30172] [Medline: 27332968]
27. Ryhänen AM, Rankinen S, Siekkinen M, Saarinen M, Korvenranta H, Leino-Kilpi H. The impact of an empowering internet-based breast cancer patient pathway program on breast cancer patients' clinical outcomes: a randomised controlled trial. J Clin Nurs 2013 Apr;22(7-8):1016-1025. [doi: 10.1111/jocn.12007] [Medline: 23480498]
28. Steel JL, Geller DA, Kim KH, Butterfield LH, Spring M, Grady J, et al. Web-based collaborative care intervention to manage cancer-related symptoms in the palliative care setting. Cancer 2016 Apr 15;122(8):1270-1282 [FREE Full text] [doi: 10.1002/cncr.29906] [Medline: 26970434]
29. van den Berg SW, Gielissen MFM, Custers JAE, van der Graaf TA, Ottevanger PB, Prins JB. BREATH: Web-based self-management for psychological adjustment after primary breast cancer--results of a multicenter randomized controlled trial. J Clin Oncol 2015 Sep 1;33(25):2763-2771. [doi: 10.1200/JCO.2013.54.9386] [Medline: 26169621]
30. van der Hout A, van Uden-Kraan CF, Holtmaat K, Jansen F, Lissenberg-Witte BI, Nieuwenhuijzen GA, et al. Role of eHealth application Oncokompas in supporting self-management of symptoms and health-related quality of life in cancer survivors: a randomised, controlled trial. Lancet Oncol 2020 Jan;21(1):80-94. [doi: 10.1016/S1470-2045(19)30675-8] [Medline: 31838009]
31. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010 Jun 1;152(11):726-732. [doi: 10.7326/0003-4819-152-11-201006010-00232] [Medline: 20335313]
32. Lee MK, Yun YH, Park H, Lee ES, Jung KH, Noh D. A web-based self-management exercise and diet intervention for breast cancer survivors: pilot randomized controlled trial. Int J Nurs Stud 2014 Dec;51(12):1557-1567. [doi: 10.1016/j.ijnurstu.2014.04.012] [Medline: 24856854]
33. Berry DL, Hong F, Halpenny B, Partridge AH, Fann JR, Wolpin S, et al. Electronic self-report assessment for cancer and self-care support: results of a multicenter randomized trial. J Clin Oncol 2014 Jan 20;32(3):199-205 [FREE Full text] [doi: 10.1200/JCO.2013.48.6662] [Medline: 24344222]
34. Foley NM, O'Connell EP, Lehane EA, Livingstone V, Maher B, Kaimkhani S, et al. PATI: Patient accessed tailored information: a pilot study to evaluate the effect on preoperative breast cancer patients of information delivered via a mobile application. Breast 2016 Dec;30:54-58. [doi: 10.1016/j.breast.2016.08.012] [Medline: 27611236]

35. Urech C, Grossert A, Alder J, Scherer S, Handschin B, Kasenda B, et al. Web-based stress management for newly diagnosed patients with cancer (STREAM): a randomized, wait-list controlled intervention study. J Clin Oncol 2018 Mar 10;36(8):780-788 [FREE Full text] [doi: 10.1200/JCO.2017.74.8491] [Medline: 29369731]
36. Gustafson DH, DuBenske LL, Namkoong K, Hawkins R, Chih M, Atwood AK, et al. An eHealth system supporting palliative care for patients with non-small cell lung cancer: a randomized trial. Cancer 2013 May 1;119(9):1744-1751 [FREE Full text] [doi: 10.1002/cncr.27939] [Medline: 23355273]
37. Bantum EO, Albright CL, White KK, Berenberg JL, Layi G, Ritter PL, et al. Surviving and thriving with cancer using a web-based health behavior change intervention: randomized controlled trial. J Med Internet Res 2014;16(2):e54 [FREE Full text] [doi: 10.2196/jmir.3020] [Medline: 24566820]
38. Bouma G, Admiraal JM, de Vries EG, Schröder CP, Walenkamp AM, Reyners AK. Internet-based support programs to alleviate psychosocial and physical symptoms in cancer patients: a literature analysis. Crit Rev Oncol Hematol 2015 Jul;95(1):26-37. [doi: 10.1016/j.critrevonc.2015.01.011] [Medline: 25701515]
39. Knowles SR, Mikocka-Walus A. Utilization and efficacy of internet-based eHealth technology in gastroenterology: a systematic review. Scand J Gastroenterol 2014 Apr;49(4):387-408. [doi: 10.3109/00365521.2013.865259] [Medline: 24494974]
40. Ramadas A, Quek KF, Chan CK, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]
41. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
42. Husson O, Mols F, van de Poll-Franse LV. The relation between information provision and health-related quality of life, anxiety and depression among cancer survivors: a systematic review. Ann Oncol 2011 Apr;22(4):761-772 [FREE Full text] [doi: 10.1093/annonc/mdq413] [Medline: 20870912]
43. Ventura F, Ohlén J, Koinberg I. An integrative review of supportive e-health programs in cancer care. Eur J Oncol Nurs 2013 Aug;17(4):498-507. [doi: 10.1016/j.ejon.2012.10.007] [Medline: 23158437]
44. Slev VN, Mistiaen P, Pasman HR, Verdonck-de LI, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. Int J Med Inform 2016 Mar;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DSMST: digital self-management support tool
HADS: Hospital Anxiety and Depression Scale
PA: physical activity
RCT: randomized controlled trial

Edited by R Kukafka; submitted 30.05.20; peer-reviewed by E Das, P Schofield; comments to author 30.07.20; revised version received 24.09.20; accepted 24.05.21; published 29.06.21.

Please cite as:
Adriaans DJM, Dierick-van Daele ATM, van Bakel MJHM, Nieuwenhuijzen GAP, Teijink JAW, Heesakkers FFBM, van Laarhoven HWM
Digital Self-Management Support Tools in the Care Plan of Patients With Cancer: Review of Randomized Controlled Trials
J Med Internet Res 2021;23(6):e20861
URL: https://www.jmir.org/2021/6/e20861
doi:10.2196/20861
PMID:34184997

Viewpoint

Goldilocks and the Three Bears: A Just-Right Hybrid Model to Synthesize the Growing Landscape of Publicly Available Health-Related Mobile Apps

Nancy Lau[1,2,3], PhD; Alison O'Daffer[1], BA; Joyce Yi-Frazier[1], PhD; Abby R Rosenberg[1,3,4], MD, MSc, MA

[1]Palliative Care and Resilience Lab, Center for Clinical and Translational Research, Seattle Children’s Research Institute, Seattle, WA, United States
[2]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States
[3]Cambia Palliative Care Center of Excellence, University of Washington, Seattle, WA, United States
[4]Department of Pediatrics, University of Washington School of Medicine, WA, United States

Corresponding Author:
Nancy Lau, PhD
Palliative Care and Resilience Lab
Center for Clinical and Translational Research
Seattle Children’s Research Institute
1920 Terry Ave
Seattle, WA, 98101
United States
Phone: 1 206 884 0569
Email: nancy.lau@seattlechildrens.org

Abstract

Mobile health (mHealth) technologies have provided an innovative platform for the deployment of health care diagnostics, symptom monitoring, and prevention and intervention programs. Such health-related smartphone apps are universally accepted by patients and providers with over 50 million users worldwide. Despite the rise in popularity and accessibility among consumers, the evidence base in support of health-related apps has fallen well behind the rapid pace of industry development. To bridge this evidence gap, researchers are beginning to consider how to best apply evidence-based research standards to the systematic synthesis of the mHealth consumer market. In this viewpoint, we argue for the adoption of a “hybrid model” that combines a traditional systematic review with a systematic search of mobile app download platforms for health sciences researchers interested in synthesizing the state of the science of consumer apps. This approach, which we have successfully executed in a recent review, maximizes the benefits of traditional and novel approaches to address the essential question of whether popular consumer mHealth apps work.

(J Med Internet Res 2021;23(6):e27105) doi:10.2196/27105

KEYWORDS

telemedicine; smartphone; mobile phones; mHealth; mobile apps; health services

Introduction

In the past decade, smartphones have become ubiquitous across personal, social, and vocational domains [1], regardless of gender, race, ethnicity, and socioeconomic status [2]. There are 3.5 billion smartphone users worldwide [3]. Nearly 75% of Americans own a smartphone and 83% of smartphone owners never leave home without it [1,4]. Mobile health (mHealth) technologies may improve access to health care by overcoming financial constraints and geographical barriers; 73% of families living below the poverty line have 1 or more smartphones even if they lack access to other resources, and telehealth enables expanded access to services in rural communities [5,6]. Over 50 million people use apps for health monitoring and diagnostic purposes worldwide [7]. Smartphone app-based tools for diagnosis, symptom monitoring, behavioral change, provider–patient communication, and disease-related education have become increasingly popular and have the potential to improve health and behavioral outcomes [8-10].

Previous research suggests that both patients and providers have a strong interest in utilizing mHealth technologies as part of health care practices, particularly apps that are supported by research evidence [11-13]. However, the research evidence lags far behind the exponential growth of publicly available apps for consumer download. As of 2020, there were approximately 101,000 mHealth apps available in major app stores with 3.7 billion annual downloads, and the market is forecasted to reach US $312 billion by 2027 [14-17]. This presents a significant research–practice divide with the widespread adoption of app-based health care tools and interventions that may not be backed by science [18].

Viewpoint Structure

This viewpoint is structured according to the main points of the “design science” framework for information systems research: problem identification, objectives, design and development, and demonstration [19].

Problem

Further fracturing the research–practice divide into a chasm, there are no gold-standard methods for evaluating the evidence in support of the efficacy of publicly available consumer apps within the growing mHealth industry.

Objectives

To address the research–practice chasm, we endeavored to develop methodologically rigorous and reproducible standards for evaluating whether publicly available mHealth tools and interventions work. In this viewpoint, our primary objective is to provide a narrative description of the lessons we learned from the process of designing a recent study evaluating the evidence in support of popular stress management and psychosocial wellness apps [20].

Design and Development

We considered methodological approaches such as a traditional systematic review and a novel search of mobile app download platforms. A systematic review is particularly well-suited for research questions pertaining to feasibility and efficacy of apps developed in research settings. Novel searches of mobile app download platforms allow researchers to examine the functionality and usability of popular apps. Neither was sufficient alone for novel research questions pertaining to the state of the science of popular apps available for consumer download. Ultimately, we decided on a “hybrid model” combining a systematic review with a systematic search of mobile app download platforms, a methodological approach that was “not too hot, not too cold, but just right.” We recommend that researchers interested in the review and synthesis of publicly available consumer mHealth apps in their respective disciplines utilize a “hybrid model” such as this to guide research design conceptualization.

Demonstration

We present 2 illustrative examples of studies following a “hybrid model” design, while providing additional citations of other successful studies.

Porridge Bowl #1: Systematic Review

First, we explored the possibility of utilizing a systematic review to synthesize the consumer app landscape. This traditional approach benefits from gold standards that have been extensively detailed in references such as the *Cochrane Handbook for Systematic Reviews of Interventions* [21] which covers all aspects of review planning from idea inception to data collection and analysis. Health science researchers interested in summarizing the state of the science in mHealth-related topics have conducted traditional systematic reviews utilizing databases of references such as Ovid MEDLINE, Embase, Cochrane Central Register of Controlled Trials, Web of Science, Scopus, and PsycINFO using the PICO framework to inform the search, where prespecified parameters include Patient problem/Population, Intervention, Comparator, and Outcomes of interest [22-31]. Previous smartphone app systematic reviews have spanned topics such as health behavior change interventions, medication management, and cognitive behavioral therapy and behavioral activation apps for depression [22-30].

Despite the advantage of rigorous well-defined methods with reproducible results, there are fatal flaws in the application of this approach specifically for those interested in asking and answering research questions pertaining to the state of the science of the mHealth industry space. First, the majority of apps developed in traditional laboratory-based and research settings are not available for public download and require private access codes provided to research participants only [26]. Second, less than 1% of mHealth apps across a range of previously explored health domains had corresponding scientific publications describing their efficacy [20,24,32]. Taken together, there is little to no overlap between apps evaluated in traditional systematic reviews (which query the existing literature in extensive library databases) and the types of consumer apps available for public download (which query the existing apps available for your mobile devices). For example, a systematic review of mHealth psychological interventions for anxiety which showed small to medium effect sizes can only tell us about the efficacy of apps that have been formally tested in clinical trials [26]; it does not provide information regarding the efficacy of the types of apps we all download directly onto our smartphones based on popularity metrics such as Top 100 lists, Editor’s Picks, media buzz, and consumer ratings. Apps developed in industry and research settings are siloed tracks. Thus, conclusions drawn from traditional systematic reviews are limited for informing the types of novel research questions endeavoring to synthesize the landscape of 101,000+ mHealth apps currently available to smartphone users.

Porridge Bowl #2: Systematic Search of Mobile App Platforms

Second, we explored the possibility of utilizing a systematic search of mobile app platforms to synthesize the state of the science of the mHealth industry. In the past decade, mHealth researchers have conducted mobile app download platform searches as an alternative method to traditional systematic reviews befitting the consumer app space [23,33-40]. (Editorial note: Note that JMIR journals do not apply the term "systematic review" to these kinds of studies, but calls them "Systematic Searches on App Stores" or similar; the term "systematic review" is reserved for literature reviews. Other publishers/journals may not distinguish these different study types). This user-centered approach prioritizes broad applicability of findings to day-to-day mobile phone users seeking digital health-management tools and interventions. Mobile app download platforms are utilized as the equivalent of library databases for data extraction in order to identify, screen, and review apps for inclusion and exclusion. In addition, the Mobile Application Rating Scale (MARS) is a commonly used tool for assessing the quality of mHealth apps and provides objective classifications (eg, price, platform, aspects of health targeted), and subjective subscale ratings in the domains of engagement, functionality, aesthetics, and information quality along with a composite app quality rating [41]. The 23-item scale has demonstrated high internal consistency and fair interrater reliability with independent coders [41]. Thus, researchers may compare and gauge the potential impact or value of mHealth programs that happen to be available for consumer download. In the United States, 54.4% of smartphone owners use Android devices and 44.3% use Apple devices [42]. Thus, the majority of existing systematic app searches (and thus app content and quality assessments) span Apple and Android platforms only [33,34,38-40]. Previous systematic searches of mobile app download platforms have spanned a wide range of topics such as smoking cessation, mindfulness, physical activity promotion, and pharmacology education [33,35,36,38,40].

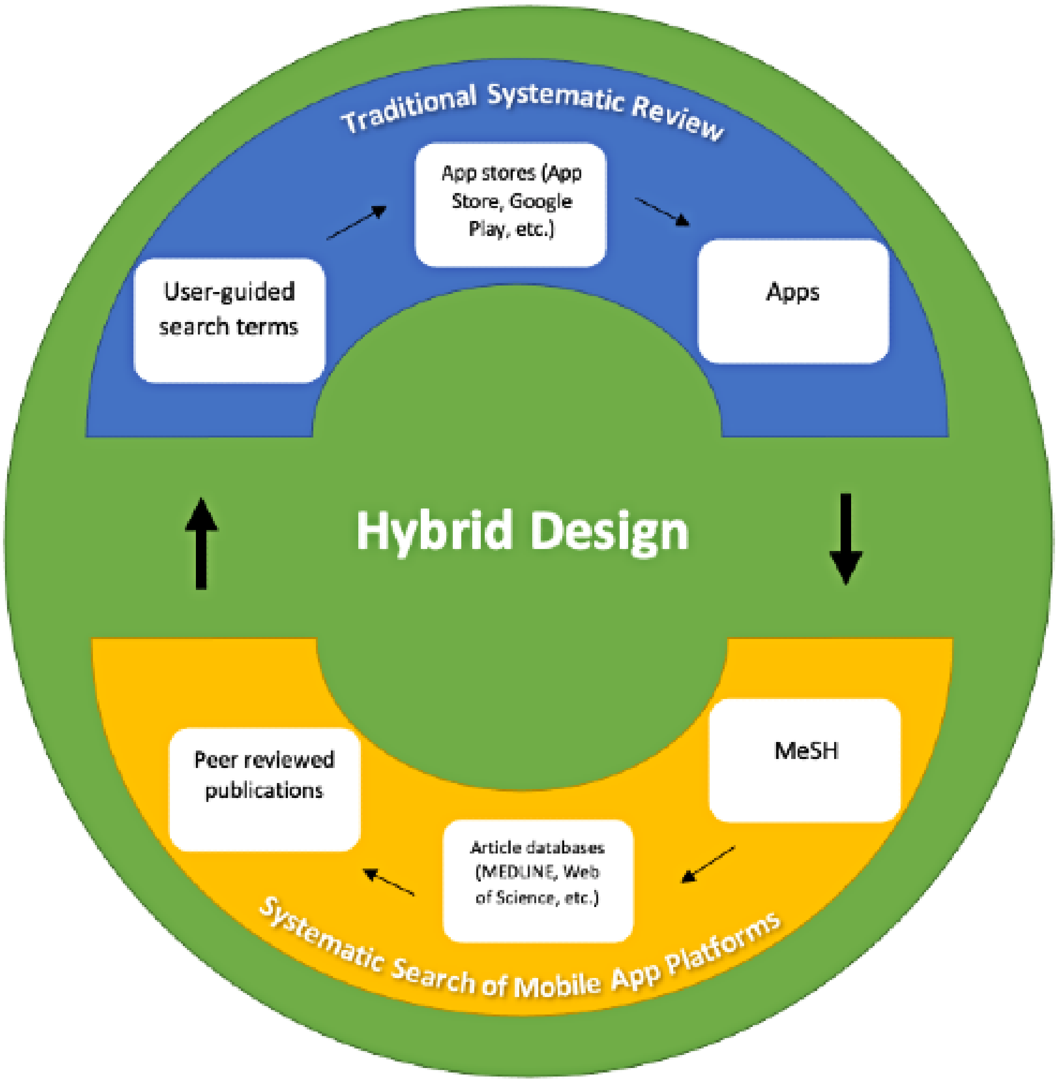
Recent methodological innovations are as follows: (1) The European Innovation and Knowledge mHealth Hub is a project established in 2020 by the International Telecommunication Union (ITU) in collaboration with the World Health Organization (WHO) [43]. The mHealth Hub offers an overview of 24 health app assessment frameworks evaluating domains including privacy, transparency, safety, and technical stability; this provides additional resources and tools for researchers to systematically synthesize app features and content. (2) Big data innovations have focused on developing automated methods to extract information on app features and components from the web using natural language processing and text analytics [44].

Despite the advantages of “real-world” representativeness of apps, research questions are limited to a synthesis of user-centered metrics such as mHealth app usability, functionality, engagement, consumer appeal, and content [45-47]. Although such scientific inquiries are important, a systematic search of mobile app platforms does not provide information on whether and how an app works, which are the scientific merits that providers and researchers rely on for establishing evidence-based standards of care and treatment recommendations. In addition, FDA-approved apps that are validated medical devices “for diagnosis of disease or other conditions, or the cure, mitigation, treatment, or prevention of a disease” are not separately listed from nonvalidated health apps; this further limits the ability of users to make informed choices about which health apps are certified tools subject to regulatory oversight [48,49].

Porridge Bowl #3: “Just Right” Hybrid Designs

To recap, systematic review methodology confers the benefit of providing information on the scientific merit of apps developed in research settings but do not represent the “real-world” consumer apps that we all download to our smartphones. App download platform search methodology, by contrast, confers the benefit of “real-world” app quality, content, and representation but not of scientific merit. Recent research has utilized hybrid design methodology by combining traditional literature review methods with systematic searches of mobile app platforms, bridging well-established traditional and novel methodologies (Figure 1) [20,24,25,32]. Next, we describe 2 illustrative examples.

Figure 1. Hybrid design model.



Illustrative Example #1: Systematic App Search Followed by Literature Review

We utilized a hybrid design to evaluate the following research aims: (1) What are the therapeutic contents and features of popular stress management and psychosocial wellness apps, and (2) Which apps, if any, are supported by peer-reviewed original research publications? This study was conducted in 3 stages [20].

Step 1 (User-Centered Approach)

We identified conventional self-help search terms from the background literature and refined the list in consensus conversations among our interdisciplinary team of intervention science researchers, health services researchers, physicians, social workers, and psychologists. Then, we input search terms directly into Android and Apple iOS mobile app search engines to identify consumer apps meeting inclusion criteria. Two authors (NL and AO) independently reviewed all apps. Independent raters created a comprehensive database with content categories representing all intervention ingredients identified, and abstracted relevant data from product pages.

Step 2 (Traditional Review Approach)

A literature review was conducted via Google Scholar, Medline, and PsycINFO databases of all commercially available apps identified in Step 1 using the search terms "[app name]" AND smartphone. Data on feasibility and efficacy outcomes were abstracted from the journal articles, and risk of bias was coded by 2 independent raters (NL and AO).

Step 3 (Synthesize Findings)

Using this hybrid design, we answered these complementary questions relevant to the state of the science of the mHealth market: Which everyday consumer apps are popular and what is their treatment content? What is the evidence in support of everyday consumer apps available to the general public?

Illustrative Example #2: Cyclical Approach

de la Vega and colleagues [32] conducted a systematic review of pain-related apps for pain symptom assessment and education. Utilizing a cyclical model, they conducted independent parallel searches, in one case starting with a systematic search of mobile app download platforms followed by a traditional literature review, and in the other case starting with a traditional literature review followed by a systematic search of mobile app download platforms.

Search 1/Step 1 (Traditional Review Approach)

A literature review was conducted across 17 relevant scientific databases (Medline, PsycINFO, Web of Science, etc.) utilizing Boolean-operator pain “AND” mobile app search terms.

Search 1/Step 2 (User-Centered Approach)

Apps meeting inclusion criteria identified in Search 1/Step 1 were input directly into 5 mobile app download platforms (Apple Store, Google Play, Blackberry App World, Nokia Store, and Windows Play Store) to determine if research-supported apps were available for public download.

Search 2/Step 1 (User-Centered Approach)

User-friendly search terms related to pain management were input directly into mobile app download platforms (Apple Store, Google Play, Blackberry App World, Nokia Store, and Windows Play Store).

Search 2/Step 2 (Traditional Review Approach)

A literature review was conducted via the aforementioned library databases of all commercially available apps identified in Search 2/Step 1 by app name. Authors also searched Google and MyHealthApps.net to learn about app origins and creators.

Step 3 (Synthesize Findings)

Using this hybrid design, de la Vega and colleagues [32] answered these complementary questions relevant to the state of the science of the mHealth market: Which research-supported apps identified for inclusion via a traditional literature review are commercially available? How many commercially available apps identified for inclusion via a systematic search of app download platforms are supported by peer-reviewed publications?

Strengths of the “hybrid model” included the cumulative advantages of systematic review and systematic search methodologies, and a comprehensive holistic analysis of the subset of “real-world” consumer apps that are research based. A hybrid design was the only approach that would allow researchers, clinicians, and patients/consumers alike to answer the essential question of whether consumer app-based health care tools and interventions that have been increasingly adopted worldwide actually work and what therapeutic content and features are incorporated in their design. In our previous study, we found further evidence to support the conclusion that apps developed in industry and research settings are siloed (ie, consumer apps developed in the mHealth industry and available for public download rarely have corresponding research publications). The “hybrid” approach allows health sciences researchers to identify the subset of apps that are both research based and publicly available despite the fact that there is no existing database/repository for consumers seeking evidence-based care.

Discussion

Smartphone apps to address a diverse array of health care needs are being developed at a rapid rate and are widely adopted worldwide. However, the scientific merit of “real-world” apps remains largely understudied and unknown. This is due, in part, to the absence of well-established methods for the evaluation of the efficacy of consumer apps. Of the 2 more common approaches to evaluation, neither is “just right” to determine both evidence base and quality for the consumer app space. Traditional literature reviews are valuable for gathering and synthesizing information regarding the scientific backing of popular apps in the form of feasibility and efficacy study data. Review of mobile app search engines allows for a direct synthesis of popular consumer apps and user-centered metrics such as usability, engagement, functionality, and app content. The “hybrid model” described in this viewpoint allows researchers to address novel research questions leveraging the complementary strengths of a systematic review and app search engine review. Although not without its limitations, hybrid approaches provide a unique opportunity to develop and iteratively refine methodologies for synthesizing the state of the science of the quickly evolving consumer mHealth market. Future research should focus on the standardization of mobile app download platform searches and systematic blending of traditional and novel methodological approaches via “just-right” hybrid designs. Such research endeavors will help bridge the research–practice chasm by rigorously evaluating digital health industry solutions to health care problems. Ultimately, this will help us understand whether popular apps work and inform mHealth clinical practice guidelines.

Acknowledgments

NL is funded as an Implementation Science Scholar through the National Heart, Lung, and Blood Institute of the National Institutes of Health (Grant number 5K12 HL137940-02). The opinions herein represent those of the authors and not necessarily the funders.

Conflicts of Interest

None declared.

References

1. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]

2. Anderson M, Jiang J. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2021-05-26]
3. O'Dea S. Number of smartphone users worldwide from 2016 to 2021 Published 2020. 2020. URL: https://www.statista.com [accessed 2021-05-26]
4. Mobile FS. Mobile Fact Sheet. Pew Research Center. 2018. URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2021-05-26]
5. Rideout V, Katz V. Opportunity for All? Technology and Learning in Lower-Income Families. 2016. URL: https://www.joanganzcooneycenter.org/wp-content/uploads/2016/01/jgcc_opportunityforall.pdf [accessed 2021-05-26]
6. Myers CR. Using Telehealth to Remediate Rural Mental Health and Healthcare Disparities. Issues Ment Health Nurs 2019 Mar;40(3):233-239. [doi: 10.1080/01612840.2018.1499157] [Medline: 30508400]
7. Millenson ML, Baldwin JL, Zipperer L, Singh H. Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis. Diagnosis (Berl) 2018 Dec 25;5(3):95-105. [doi: 10.1515/dx-2018-0009] [Medline: 30032130]
8. Torous J, Chan SR, Yee-Marie TS, Behrens J, Mathew I, Conrad EJ, et al. Patient Smartphone Ownership and Interest in Mobile Apps to Monitor Symptoms of Mental Health Conditions: A Survey in Four Geographically Distinct Psychiatric Clinics. JMIR Ment Health 2014;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
9. Featherall J, Lapin B, Chaitoff A, Havele SA, Thompson N, Katzan I. Characterization of Patient Interest in Provider-Based Consumer Health Information Technology: Survey Study. J Med Internet Res 2018 Apr 19;20(4):e128 [FREE Full text] [doi: 10.2196/jmir.7766] [Medline: 29674312]
10. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
11. Apolinário-Hagen J, Hennemann S, Fritsche L, Drüge M, Breil B. Determinant Factors of Public Acceptance of Stress Management Apps: Survey Study. JMIR Ment Health 2019 Nov 07;6(11):e15373 [FREE Full text] [doi: 10.2196/15373] [Medline: 31697243]
12. Schueller SM, Washburn JJ, Price M. Exploring Mental Health Providers' Interest in Using Web and Mobile-Based Tools in their Practices. Internet Interv 2016 May;4(2):145-151. [doi: 10.1016/j.invent.2016.06.004] [Medline: 28090438]
13. Berkowitz CM, Zullig LL, Koontz BF, Smith SK. Prescribing an App? Oncology Providers' Views on Mobile Health Apps for Cancer Care. JCO Clin Cancer Inform 2017 Dec;1:1-7 [FREE Full text] [doi: 10.1200/CCI.17.00107] [Medline: 30657404]
14. Globe Newswire. Mobile Health (mHealth) Market To Reach USD 311.98 Billion By 2027. 2020 Apr 28. URL: https://tinyurl.com/yjr62d8d [accessed 2021-05-26]
15. Pohl M. mHealth App Economics 2017/2018: Current Status and Future Trends in Mobile Health. 2017. URL: https://research2guidance.com/wp-content/uploads/2017/11/R2G-mHealth-DeveloperEconomics-2017-Status-And-Trends.pdf [accessed 2021-05-26]
16. Statista. Number of mHealth Apps Available at Google Play From 1st Quarter 2015 to 4th Quarter 2020. 2021. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2021-05-26]
17. Statista. Number of mHealth Apps Available in the Apple App Store from 1st Quarter 2015 to 4th Quarter 2020. 2021. URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ [accessed 2021-04-26]
18. Aitken M. Patient Adoption of mHealth: Use, Evidence, and Remaining Barriers to Mainstream Acceptance. Minneapolis, MN: Institute for Health Informatics; 2015.
19. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 2014 Dec 08;24(3):45-77. [doi: 10.2753/MIS0742-1222240302]
20. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone Mobile Apps for Psychosocial Wellness and Stress Management: Systematic Search in App Stores and Literature Review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
21. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. London, UK: The Cochrane Collaboration and John Wiley & Sons; 2019.
22. Chib A, Lin SH. Theoretical Advancements in mHealth: A Systematic Review of Mobile Apps. J Health Commun 2018;23(10-11):909-955 [FREE Full text] [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
23. Franko OI. Smartphone apps for orthopaedic surgeons. Clin Orthop Relat Res 2011 Jul;469(7):2042-2048 [FREE Full text] [doi: 10.1007/s11999-011-1904-0] [Medline: 21547414]
24. MacKenzie Greenle M, Morgan B, Sayani S, Meghani SH. Identifying Mobile Apps Targeting Palliative Care Patients and Family Members. J Palliat Med 2018 Oct;21(10):1380-1385. [doi: 10.1089/jpm.2018.0157] [Medline: 30312126]
25. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M, Herreros-González J. Mobile apps in cardiology: review. JMIR Mhealth Uhealth 2013;1(2):e15 [FREE Full text] [doi: 10.2196/mhealth.2737] [Medline: 25098320]
26. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]

27. Han M, Lee E. Effectiveness of Mobile Health Application Use to Improve Health Behavior Changes: A Systematic Review of Randomized Controlled Trials. Healthc Inform Res 2018 Jul;24(3):207-226. [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
28. Hieftje K, Edelman EJ, Camenga DR, Fiellin LE. Electronic media-based health interventions promoting behavior change in youth: a systematic review. JAMA Pediatr 2013 Jun;167(6):574-580 [FREE Full text] [doi: 10.1001/jamapediatrics.2013.1095] [Medline: 23568703]
29. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A Systematic Review of Cognitive Behavioral Therapy and Behavioral Activation Apps for Depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
30. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
31. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007;7:16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
32. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
33. Abroms LC, Lee WJ, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
34. Blázquez Martín D, De La Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and Controlling Stress Using mHealth: Systematic Search in App Stores. JMIR Mhealth Uhealth 2018 May 09;6(5):e111 [FREE Full text] [doi: 10.2196/mhealth.8866] [Medline: 29743152]
35. Coulon SM, Monroe CM, West DS. A Systematic, Multi-domain Review of Mobile Smartphone Apps for Evidence-Based Stress Management. Am J Prev Med 2016 Jul;51(1):95-105. [doi: 10.1016/j.amepre.2016.01.026] [Medline: 26993534]
36. Haffey F, Brady RRW, Maxwell S. Smartphone apps to support hospital prescribing and pharmacology education: a review of current provision. Br J Clin Pharmacol 2014 Jan;77(1):31-38. [doi: 10.1111/bcp.12112] [Medline: 23488599]
37. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and Evaluation of Mindfulness-Based iPhone Apps. JMIR Mhealth Uhealth 2015;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
38. Middelweerd A, Mollee JS, van der Wal CN, Brug J, Te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
39. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. J Med Internet Res 2015;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
40. Weekly T, Walker N, Beck J, Akers S, Weaver M. A Review of Apps for Calming, Relaxation, and Mindfulness Interventions for Pediatric Palliative Care Patients. Children (Basel) 2018 Jan 26;5(2):16 [FREE Full text] [doi: 10.3390/children5020016] [Medline: 29373515]
41. 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;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
42. Statista. Subscriber Share Held by Smartphone Operating Systems in the United States from 2012 to 2018. 2019. URL: https://www.statista.com/statistics/266572/market-share-held-by-smartphone-platforms-in-the-united-states/ [accessed 2021-05-26]
43. Statistica. Knowledge Tool 1: Assessment Frameworks in mHealth. 2020. URL: https://mhealth-hub.org/download/d2-1-knowledge-tool-1-health-apps-assessment-frameworks-pending-ec-approval [accessed 2021-05-26]
44. Paglialonga A, Schiavo M, Caiani EG. Automated Characterization of Mobile Health Apps' Features by Extracting Information From the Web: An Exploratory Study. Am J Audiol 2018 Nov 19;27(3S):482-492 [FREE Full text] [doi: 10.1044/2018_AJA-IMIA3-18-0008] [Medline: 30452752]
45. Romero RL, Kates F, Hart M, Ojeda A, Meirom I, Hardy S. Quality of Deaf and Hard-of-Hearing Mobile Apps: Evaluation Using the Mobile App Rating Scale (MARS) With Additional Criteria From a Content Expert. JMIR Mhealth Uhealth 2019 Oct 30;7(10):e14198 [FREE Full text] [doi: 10.2196/14198] [Medline: 31670695]
46. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile Apps for Blood Pressure Monitoring: Systematic Search in App Stores and Content Analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
47. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer Mobile Apps for Potential Drug-Drug Interaction Check: Systematic Review and Content Analysis Using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]

48. Athanas AJ, McCorrison JM, Smalley S, Price J, Grady J, Campistron J, et al. Association Between Improvement in Baseline Mood and Long-Term Use of a Mindfulness and Meditation App: Observational Study. JMIR Ment Health 2019 May 08;6(5):e12617 [FREE Full text] [doi: 10.2196/12617] [Medline: 31066704]
49. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]

Abbreviations

ITU: International Telecommunication Union
MARS: Mobile Application Rating Scale
mHealth: mobile health
WHO: World Health Organization

Edited by R Kukafka; submitted 11.01.21; peer-reviewed by E Caiani, O El-Gayar; comments to author 17.03.21; revised version received 09.04.21; accepted 06.05.21; published 07.06.21.

Please cite as:
Lau N, O'Daffer A, Yi-Frazier J, Rosenberg AR
Goldilocks and the Three Bears: A Just-Right Hybrid Model to Synthesize the Growing Landscape of Publicly Available Health-Related Mobile Apps
J Med Internet Res 2021;23(6):e27105
URL: https://www.jmir.org/2021/6/e27105
doi:10.2196/27105
PMID:34096868

Viewpoint

Team Science in Precision Medicine: Study of Coleadership and Coauthorship Across Health Organizations

Ning An[1*], PhD; John Mattison[2*], MD; Xinyu Chen[3*], MSc; Gil Alterovitz[4*], PhD

[1]Key Laboratory of Knowledge Engineering with Big Data of the Ministry of Education, Intelligent Interconnected Systems Laboratory of Anhui Province, School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China

[2]Kaiser Permanente, Arsenal Capital Partners, New York City, NY, United States

[3]Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA, United States

[4]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Gil Alterovitz, PhD
Brigham and Women's Hospital
Harvard Medical School
75 Francis Street
Boston, MA, 02115
United States
Phone: 1 617 329 1445
Email: gil_alterovitz@hms.harvard.edu

Abstract

Background: Interdisciplinary collaborations bring lots of benefits to researchers in multiple areas, including precision medicine.

Objective: This viewpoint aims at studying how cross-institution team science would affect the development of precision medicine.

Methods: Publications of organizations on the eHealth Catalogue of Activities were collected in 2015 and 2017. The significance of the correlation between coleadership and coauthorship among different organizations was calculated using the Pearson chi-square test of independence. Other nonparametric tests examined whether organizations with coleaders publish more and better papers than organizations without coleaders.

Results: A total of 374 publications from 69 organizations were analyzed in 2015, and 7064 papers from 87 organizations were analyzed in 2017. Organizations with coleadership published more papers ($P<.001$, 2015 and 2017), which received higher citations ($Z=-13.547$, $P<.001$, 2017), compared to those without coleadership. Organizations with coleaders tended to publish papers together ($P<.001$, 2015 and 2017).

Conclusions: Our findings suggest that organizations in the field of precision medicine could greatly benefit from institutional-level team science. As a result, stronger collaboration is recommended.

(J Med Internet Res 2021;23(6):e17137) doi:10.2196/17137

KEYWORDS

precision medicine; team science

Introduction

The concept of a meta-topical brainforest is proposed, to reflect a link between collaborative research and complex ecosystems. Tropical rainforests leverage a diversity of species to capture and convert solar energy into carbon-based life, and research teams can harvest a similar benefit from a diversity of data, tools, and thought paradigms.

According to the National Institutes of Health, team science is “a collaborative and often cross-disciplinary approach to scientific inquiry that draws researchers who otherwise work independently or as coinvestigators on smaller-scale projects into collaborative centers and groups” [1]. Thus, team science occurs when artificial boundaries such as departments and institutions are crossed, allowing collaboration in integrated networks. Over the past two decades, the concept has received increasing attention to better understand and address global challenges [2-5]. In 2007, Wuchty et al [6] examined 19.9 million research articles in the Institute for Scientific Information Web of Science database and 2.1 million patent records on multiple topics. They concluded that a team-authored paper has increased probability of being highly cited. The systems being formed through interdisciplinary collaborations help teams reach achievements that individual researchers are less likely to accomplish.

Kohane [7] pointed out that precision medicine in particular requires a higher level of coordination between various agencies and suggests the boundaries between research projects and clinical care institutions should be blurred to link gathered data. The exponential growth and causal interdependencies of “-omics” fields dictate that expertise across disciplines is essential to making meaningful and durable contributions to the understanding of human biology.

This brief viewpoint aims to explore the impact of cross-institution team science on the development of precision medicine. We hypothesized that international organizations with coleaders tend to publish more impactful papers than organizations without coleaders. Using the Pearson chi-square test and the Mann-Whitney *U* test, we validated our hypothesis.

Methods

Information was collected from the eHealth Catalogue of Activities developed by the nonprofit Global Alliance for Genomics and Health in 2015 [8]. The catalog lists international genomic and clinical data-sharing initiatives, and the eHealth Task Team updated the catalog through 2017. The data on the executive leadership team and publications were obtained from the websites of these organizations. If such information was not found, additional data were acquired by directly contacting the organizations or searching on Google Scholar. The impact of papers was evaluated by their number of citations, a criterion of research quality [6].

In this paper, coleadership means that a person holds a leadership position in different organizations concurrently. If two papers from separate organizations have at least one author in common, these two organizations are regarded as having a coauthor relationship.

Nonparametric tests were performed to verify the hypothesis. We used SPSS (version 22.0; IBM Corp) and R to perform two-tailed tests with an α level of .05. The significance of the correlation between the nominal variables coleadership and coauthorship was examined using the Pearson chi-square test of independence and expressed in a contingency table. The Pearson chi-square test of goodness of fit was adopted to evaluate whether organizations with coleaders had a greater number of publications than organizations without coleaders, and the Mann-Whitney *U* test was used to examine whether the former organizations published papers that received more citations than the latter.

Results

Overview

We analyzed data from 69 organizations in the catalog and found 16 pairs with coleader relationships in 2015. Among the 374 publications from these organizations at that time, 13 pairs had coauthors. By 2017, the number of institutions in the catalog increased to 87, and there were 37 pairs with coleadership, corresponding to 30 organizations. Information on 7064 papers was collected, showing that 55 organizations had coauthored publications, with 436 papers in total.

Number of Publications

The chi-square goodness of fit test suggests that the number of papers being published is strongly correlated with the category of the organization—organizations in a coleadership network or organizations without coleadership ($P<.001$, 2015 and 2017).

Quality of Publications

The citation number of each paper was obtained from Google Scholar. The results of the Mann-Whitney *U* test indicated that the number of citations received by publications of organizations with and without coleaders differed significantly ($Z=–13.547$, $P<.001$, 2017). Papers from the former organizations had a higher mean rank (3603.35 for the group of papers whose authors are in the coleadership network, and 2702.67 for the other group), which means that the organizations with coleaders tended to have a greater number of highly cited papers.

Relationship Between Coleader and Coauthor

In the chi-square test of independence, the total sample size is the number of lines in a fully connected diagram. The results indicate that in both 2015 and 2017, organizations with coleaders tended to publish papers together, suggesting that coleadership will lead to coauthorship ($P<.001$, 2015 and 2017).

Discussion

We studied how precision medicine can be influenced by institutional-level team science by analyzing coleadership and coauthorship across health organizations. From 2015 to 2017, the number of health organizations grew from 69 to 87, and their publications increased. Concurrent positions held by leaders may incentivize researchers to work for multiple organizations; thus, the researchers will be very likely to have a coauthored paper ($P<.001$, both 2015 and 2017). Moreover, the publications from organizations with coleaders are more frequently cited, indicating a relatively high quality ($Z=–13.547$, $P<.001$). These results suggest that collaborations among health institutions are becoming stronger, which promotes their working efficiency.

These results illustrate the concept of meta-topical brainforests in precision medicine (Figures 1-2) and may have broader implications: cross-enterprise cooperation plays an essential role in solving complex issues. As a field-crossing example, Sovacool [9] suggested researchers should incorporate expertise and data from indigenous groups to address global environmental challenges.

One hopes the analogy persists and the extraordinary natural future-proofing mechanisms in rainforests coincide with similar continued diversification in research networks and widely impactful publications.

Figure 1. The coleader relationship network in 2017, with nodes representing organizations and lines representing concurrent coleadership.

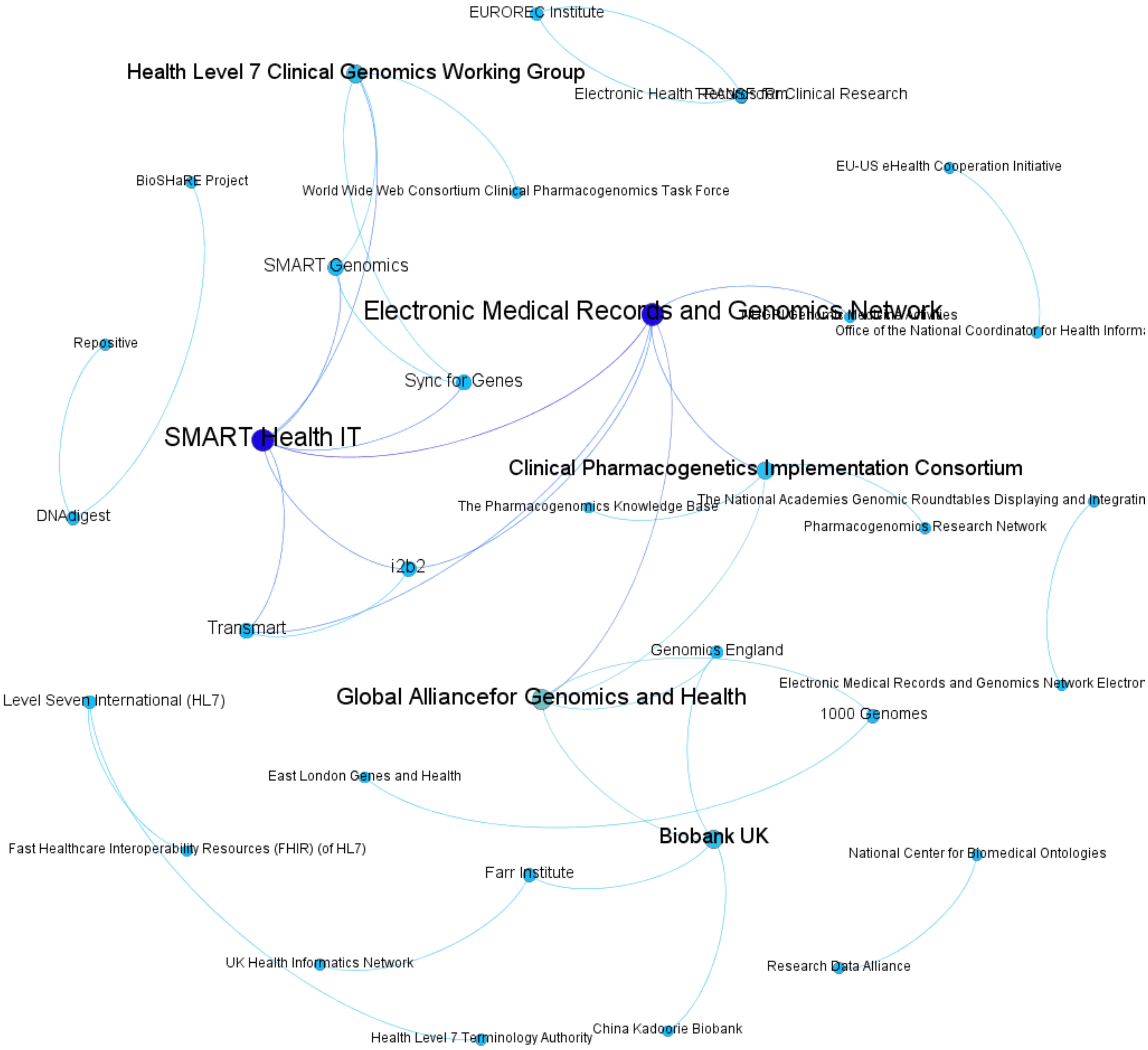
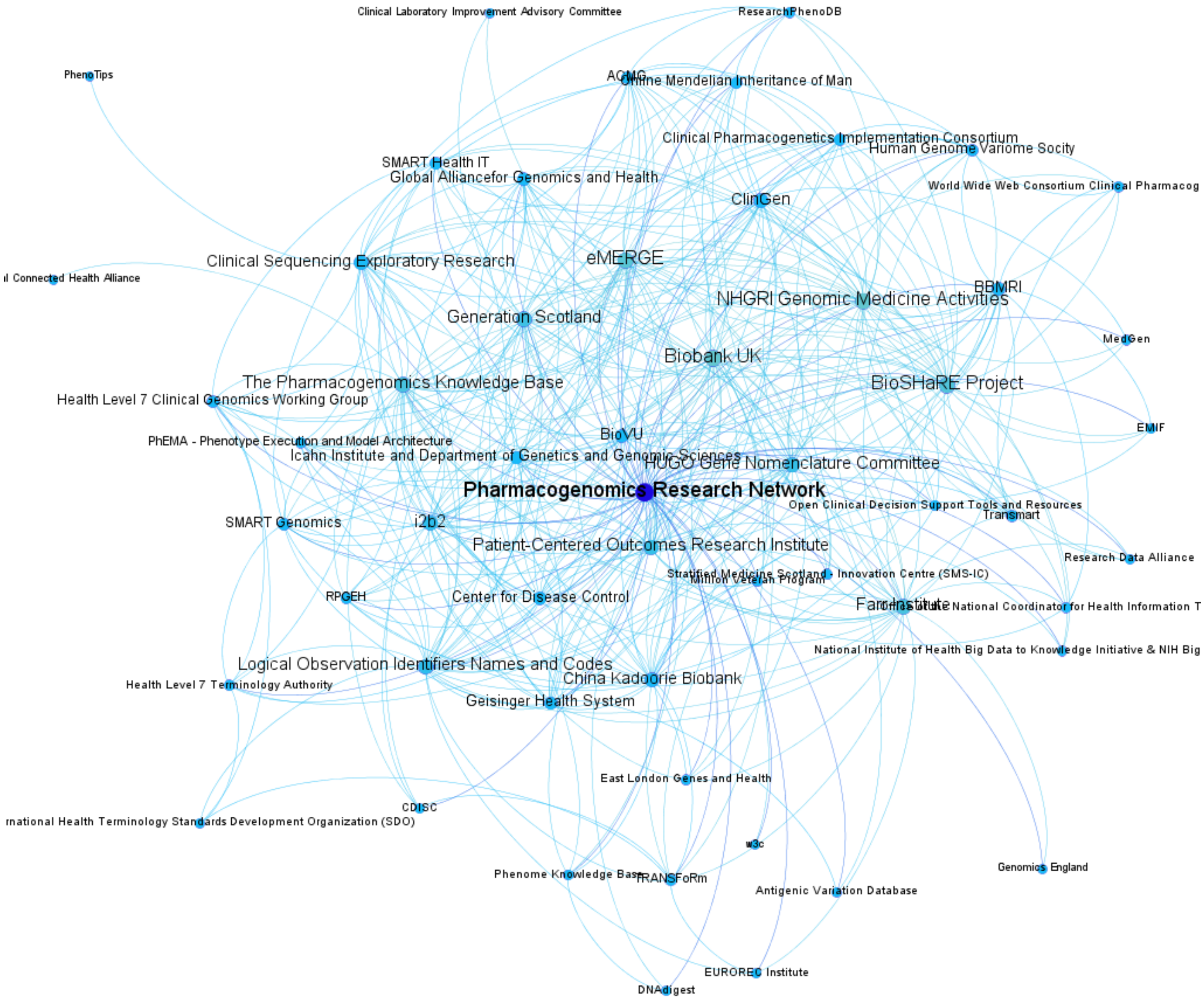


Figure 2. The coauthor relationship in 2017, with nodes representing organizations and lines connecting organizations by coauthored publications. Nodes darken with more connected lines.



Conflicts of Interest

None declared.

References

1. Bennett L, Gadlin H, Marchand C. Collaboration and Team Science Field Guide, 2nd ed.: NIH National Cancer Institute, Center for Research Strategy; 2018. URL: https://www.cancer.gov/about-nci/organization/crs/research-initiatives/team-science-field-guide/collaboration-team-science-guide.pdf [accessed 2021-05-31]
2. Stokols D, Misra S, Moser RP, Hall KL, Taylor BK. The ecology of team science: understanding contextual influences on transdisciplinary collaboration. Am J Prev Med 2008 Aug;35(2 Suppl):S96-115. [doi: 10.1016/j.amepre.2008.05.003] [Medline: 18619410]
3. Shaikh AR, Butte AJ, Schully SD, Dalton WS, Khoury MJ, Hesse BW. Collaborative biomedicine in the age of big data: the case of cancer. J Med Internet Res 2014 May 07;16(4):e101 [FREE Full text] [doi: 10.2196/jmir.2496] [Medline: 24711045]

4. Roosan D, Law AV, Karim M, Roosan M. Improving Team-Based Decision Making Using Data Analytics and Informatics: Protocol for a Collaborative Decision Support Design. JMIR Res Protoc 2019 Dec 27;8(11):e16047 [FREE Full text] [doi: 10.2196/16047] [Medline: 31774412]
5. Xu X, Hu J, Lyu X, Huang H, Cheng X. Exploring the Interdisciplinary Nature of Precision Medicine：Network Analysis and Visualization. JMIR Med Inform 2021 Jan 11;9(1):e23562 [FREE Full text] [doi: 10.2196/23562] [Medline: 33427681]
6. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. Science 2007 May 18;316(5827):1036-1039 [FREE Full text] [doi: 10.1126/science.1136099] [Medline: 17431139]
7. Kohane IS. Health care policy. Ten things we have to do to achieve precision medicine. Science 2015 Jul 03;349(6243):37-38. [doi: 10.1126/science.aab1328] [Medline: 26138968]
8. Global Alliance for Genomics and Health Summary of Third Plenary Meeting, June 2015, Leiden, the Netherlands. URL: https://www.ga4gh.org/wp-content/uploads/GlobalAlliance-3rdPlenaryMeeting.pdf [accessed 2021-05-31]
9. Sovacool BK. Diversity: Energy studies need social science. Nature 2014 Jul 31;511(7511):529-530. [doi: 10.1038/511529a] [Medline: 25079540]

Edited by G Eysenbach; submitted 22.11.19; peer-reviewed by L Monteiro, M Afzal, S Brunak; comments to author 26.10.20; revised version received 11.03.21; accepted 03.04.21; published 14.06.21.

Please cite as:
An N, Mattison J, Chen X, Alterovitz G
Team Science in Precision Medicine: Study of Coleadership and Coauthorship Across Health Organizations
J Med Internet Res 2021;23(6):e17137
URL: https://www.jmir.org/2021/6/e17137
doi:10.2196/17137
PMID:34125070

Viewpoint

A Traditional Chinese Medicine Traceability System Based on Lightweight Blockchain

Zhengfei Wang[1], DEng; Lai Wang[1], MM; Fu'an Xiao[1], MM; Qingsong Chen[2], BEng; Liming Lu[3]; Jiaming Hong[1], PhD

[1]School of Medical Information Engineering, Guangzhou University of Chinese Medicine, Guangzhou, China

[2]Shenzhen Ant Network Service Co LTD, Guangzhou, China

[3]Clinical Research and Data Center, South China Research Center for Acupuncture and Moxibustion, Medical College of Acu-Moxi and Rehabilitation, Guangzhou University of Chinese Medicine, Guangzhou, China

Corresponding Author:
Jiaming Hong, PhD
School of Medical Information Engineering
Guangzhou University of Chinese Medicine
232 Waihuandong Road, Guangzhou Higher Education Mega Center, Panyu District
Guangzhou, 510006
China
Phone: 86 13430200356
Email: hjm@gzucm.edu.cn

Abstract

Background: Recently, the problem of traditional Chinese medicine (TCM) safety has attracted attention worldwide. To prevent the spread of counterfeit drugs, it is necessary to establish a drug traceability system. A traditional drug traceability system can record the whole circulation process of drugs, from planting, production, processing, and warehousing to use by hospitals and patients. Once counterfeit drugs are found, they can be traced back to the source. However, traditional drug traceability systems have some drawbacks, such as failure to prevent tampering and facilitation of sensitive disclosure. Blockchain (including Bitcoin and Ethernet Square) is an effective technology to address the problems of traditional drug traceability systems. However, some risks impact the reliability of blockchain, such as information explosion, sensitive information leakage, and poor scalability.

Objective: To avoid the risks associated with the application of blockchain, we propose a lightweight block chain framework.

Methods: In this framework, both horizontal and vertical segmentations are performed when designing the blocks, and effective strategies are provided for both segmentations. For horizontal segmentation operations, the header and body of the blockchain are separated and stored in the blockchain, and the body is stored in the InterPlanetary File System. For vertical segmentation operations, the blockchain is cut off according to time or size. For the addition of new blocks, miners only need to copy the latest part of the blockchain and append the tail and vertical segmentation of the block through the consensus mechanism.

Results: Our framework could greatly reduce the size of the blockchain and improve the verification efficiency.

Conclusions: Experimental results have shown that the efficiency improves compared with ethernet when a new block is added to the blockchain and a search is conducted.

(J Med Internet Res 2021;23(6):e25946) doi:10.2196/25946

KEYWORDS

blockchain; traditional Chinese medicine; TCM; traceability system; fake drugs; IPFS; fraud; traceability

Introduction

In current society, drugs are an important guarantee of healthy human life. Contrastingly, counterfeit drugs cannot cure diseases and may cause physical damage to patients. In the worst cases, serious drug accidents even lead to death. According to World Trade Organization statistics, thousands of patients have died of drug accidents due to ingestion of counterfeit drugs worldwide [1]. Drug accidents are more likely to happen in developing countries, resulting in substantial medical economic losses.

For the general population and even for physicians, it is difficult to identify counterfeit drugs; owing to this difficulty, it is more difficult to prevent the use of counterfeit drugs. The circulation of traditional Chinese medicine (TCM) must go through a series of processes, from cultivation, fresh product collection, processing, and circulation of prepared pieces to their sale as TCM patent medicines [2]. Because the composition of proprietary Chinese medicines is complicated, drugs are not easily identified by the most direct senses. To effectively prevent counterfeiting of drugs and to crack down on harmful behaviors in all aspects of production, it is important to establish a reliable drug traceability system. From the planting of drugs at the base and their growing, harvesting, processing, packaging, storage, and circulation, to selling to patients, freshly collected products must go through many different processes before becoming patent medicines, such as prepared pieces. Every link of the drug circulation process should be tracked and recorded; blockchain can implement each process of every producer and seller to achieve traceable destinations and traceable sources.

Blockchain can help patients understand the past life path of a drug. Therefore, if there is a problem with the quality of the drug, it can be traced back to the source of the problem.

The United States and China have adopted the “whole process traceability” mode [3,4]. Drug traceability includes the production, circulation, use, and other aspects of drugs. This mode realizes gradual transmission and traceability of drug information and achieves interlocking links. It is conducive to preventing counterfeit and inferior drugs from entering the drug supply chain, standardizing the drug circulation order, and ensuring drug safety. The European Union has established a unified European drug certification system, which is composed of the National Drug Certification Systems of Member States, European Drug Coding Center and European central hub [5].

In May 2011, the Ministry of Commerce of China, the State Administration of Traditional Chinese Medicine, and the State Food and Drug Administration (SFDA) carried out a pilot project of a Chinese medicine traceability system in Chongqing. In 2015, the State Administration of Traditional Chinese Medicine took the lead in supporting “quality traceability,” and they proposed quality control standards for the whole process of TCM production and the standards and evaluation system of key TCM products with quality information traceability [6]. In 2018, the SFDA promoted the construction of a drug information traceability system, realized “one item, one code, and the same tracking,” strengthened the exchange and sharing of traceability information, and realized whole product and whole process traceability [7]. In 2019, the SFDA issued the *Guidelines for the Construction of Drug Information Traceability System*, *Drug Traceability Code Coding Requirements*, the *Basic Technical Requirements for Drug Traceability System*, and other informatization standards [8].

However, traceability systems pose great security risks. Because the drug production data of each link are stored in the system, these data could be easily tampered with, deleted, or denied. If such tampering were to occur, the authenticity of the traceability system would not be guaranteed, and the system would lose application value. The emergence of blockchain technology fundamentally changed this situation. Blockchain is a type of decentralized distributed account book with characteristics that cannot be tampered with, deleted, or denied. Storing traceability information in the blockchain prevents the abovementioned risks and ensures data security.

At present, blockchain has produced a new batch of technical solutions which have been implemented in clinical trials in some medical institutions [9]. Blockchain provides great convenience for the distribution management of medical data and the sharing under the control of access rights. Blockchain’s tamperproof and facile query characteristics can promote its use with medical resources.

However, these blockchain systems based on drug systems have the following problems:

1. Data explosion. As a chain record, the blocks store all the data in the drug transaction from beginning to end, and the data storage creates a heavy load on the system.
2. The low efficiency problem. The characteristics of the chained records will affect the query speed; moreover, the data capacity of blockchain is very large, which will also affect the query efficiency.
3. Security issues. Blockchain originates from Bitcoin; however, Bitcoin is not suitable for the medical field. Medical data are only disclosed to the public on the premise of ensuring security and privacy. Blockchain has distributed and multicenter characteristics. However, it is also necessary to ensure that important system data are not stolen and to prevent malicious users from attacking the data. Therefore, previous systems require improvement.

In view of the drug traceability development needs and the deficiency of blockchain applications, this paper proposes a new TCM drug traceability system based on blockchain that effectively protects data security, fundamentally realizes information traceability, and prevents data tampering and denial.

The main contributions of this work are as follows.

First, we propose a Chinese medicine traceability system architecture based on blockchain and the InterPlanetary File System (IPFS) to effectively solve the blockchain information explosion problem. In the design of the blocks, the blocks are divided horizontally, and the header and body of traditional blockchain are separated. The header is stored in the block, and the body is stored in the InterPlanetary File System (IPFS). The design achieves the effective verification of information but also reduces the block size. Further, we divided the blocks vertically. The length of the blockchain increases with the time dimension; this increase will affect the verification efficiency of the blockchain. Therefore, we selected a threshold (time or size) to segment the blocks. When users verify new blocks, they only need to copy certain parts of the blockchain.

Second, we evaluate the performance of our proposed architecture based on blockchain and the IPFS. Compared with the Ethereum network, our experiments demonstrate that our proposed architecture outperforms the Ethereum network in terms of the time cost for processing ledger update and query.

Third, in the initial stage, a blockchain is added to the original drug traceability system to ensure that the information will not be tampered with or denied. This is not a substitute but a supplement for the system. After it gradually becomes mature, the new TCM herbal medicine traceability system will be transplanted to the blockchain to maintain the smooth transition.

Related Work

Due to its decentralized and tamperproof nature, blockchain has received extensive attention in many areas, especially in food and drug traceability, as well as for electronic health record (EHR) sharing and traceability. However, there are few research results on the traceability of TCM.

Food and Drug Traceability Based on Blockchain

In 2016, Tian et al [10] developed a new blockchain system used radio-frequency identification (RFID) to improve the efficiency of collecting information automatically with a corresponding new system for storage and management. Galvez et al [11] proposed the use of blockchain to prevent economic losses and erosion of consumers' trust after viewing its potential in traceability and authenticity. Galvez et al indicated that different cases are operable, including plant food, animal food, and other industries, on the basis of existing products.

Toyoda et al [12] used smart contracts in data management and made changes to enable the smart contracts to be classified by levels; this smart contract level classification improved the management efficiency. In real-world data exploration of supply chains in IBM and Tsinghua University's collaboration work [13], IBM tested the time cost of uploading data in a public chain and showed that data exploration decreased the efficiency.

Daniel Tse et al [14] prospected a specific food supply chain blockchain in 2017; soon, San Miguel et al [15] raised concerns regarding the food supply chain, integrating agricultural information technology in traceability.

Blockchain has also been developed in drug-related areas. In Europe, a biobank was built for more comfortable data management, including medical data related to drugs. N Mamo et al [16] proposed a biobanking scheme called Dwarna for dynamic consent.

EHRs Based on Blockchain

Traditional electronic medical records (EMRs) have been gradually transformed into EHRs; this new type of digital record enables patients and physicians to manage the patient's health record together rather than allowing only the physician to control all the data, as before. This progress increased the convenience and efficiency of medical service; however, a large problem remains, namely that patients and physicians' privacy can also more readily be violated [17]. With new cloud computing technology, Dubovitskaya et al [17] used smart contract blockchain technology with new cloud computing technology for access control to guarantee high security levels.

With mobile phones, access to medical data is becoming increasingly convenient for users, including both patients and physicians; however, the tasks of uploading and sharing the data are challenging. Mobile devices can detect people's health information and share it in the cloud; however, this requires a cloud server solution that is honest-but-curious [18]. Hang et al [19] proposed a new blockchain platform for patients to access their records in a comprehensive and immutable way to meet security policies; a case study of Hyperledger Fabric and a benchmark study were used in real hospitals, and these studies revealed this platform's potential for acceleration.

In addition to the medicine protection area, C Krittanawong et al [20] introduced a blockchain application in the area of artificial intelligence; technology by Tezos used blockchain in robotic event registration and verification, and Atheon's platform used blockchain for time-resuming reduction. Retrievable information, automated uploading in lack for robots marketplace, and global transparency are provided by blockchain [20]. In some traditional work, the abovementioned problems could not be readily solved; for example, in Wu et al's [21] work, they provided total access to a cloud server but could not promise safety or authenticity of the EHRs.

Because in the work mentioned, total access to cloud severs was given, the researchers changed the initial structure of Bitcoin to avoid this weakness [22]. They used smart contracts for access control to defend against attackers. Also, smart contract and blockchain were used in data management. The architecture consisted of data offloading and data sharing through access control by verifying a public key to find the corresponding ID with the IPFS data storage system. Unlike pioneering studies, this paper describes an experiment on a proposed system with two virtual machines as miners, two as the administrator and EHR manager, IPFS on Amazon, and Rivest-Shamir-Adleman (RSA) encryption. Appropriately designed experiments afforded better performance.

Lightweight Blockchain Research

Ismail et al [23] described how a combination of distributed ledger (DL), consensus protocol, and cryptography technology can be used for blockchain updating. They proposed a blockchain-based health care data management architecture after considering different parts in which the head blockchain manager regulates the network and block transaction work mode. Changing blockchain nodes to decrease the burden, they used a method called a canal to collaborate with network participants. The architecture included a ledger; notification manager records; agreement consensus, such as the practical Byzantine fault tolerance consensus protocol, in proof of work (PoW); and data replication controlled by a strong consistency model and constant prefix consistency.

A new blockchain architecture must guarantee the safety of the blocks in the arrangement of nodes and ledgers during updating and query. In denial-of-service (DoS) and time-based attacks, it should be impossible to modify or attack the data. To address efficiency and energy cost problems, clusters for different hospital nodes with a manager maintaining the ledgers could be helpful [24]. Arjona el al [25] adopted a lightweight fingerprint recognition solution; their QFingerMap16 method can be implemented in low-cost sensor nodes using lightweight dual-factor protocols. In work by Arjona et al [25], a sensor communication protocol was implemented to defend against remote attacks and impersonation attacks.

One of the main concerns in research on the use of blockchain with health care data is effectively decreasing the update time. Lwin et al [26] built a lightweight trust system with a consensus algorithm; their scheme distributes nodes in a mobile ad hoc network to create a tamperproof and optimized link state routing protocol. Fu et al [27] used an interleaving encoding algorithm to propose a lightweight message sharing scheme. This scheme applies an interleaving encoder to conceal sensitive information. EMRs are mapped to *n* different short shares, and the shares are transmitted to different nodes on the blockchain. Fu's sharing scheme reduces the cost of shorter shares and provides an efficient reconstruction process; moreover, the use of indexes increases the system stability [27].

Current Medical Blockchain Products

Most well-known medical blockchain products use advantages of blockchain such as decentralization, immutability, security, and transparency, including solutions for third-party trust, prevention of data tampering or deletion, cryptography, data management, and data tracking and storage problems.

The most well-known blockchain products are MedRec [28], MedicoHealth [29], MeFy [30], and MediBloc [31]. Based on Bitcoin, the medical decentralized distributed ledger technology MedRec was proposed as a solution for system interoperability and provided access for more clients to manage their medical records, such as patients and physicians, with smart contract and consensus. Three types of contracts are included in MedRec for real-time sharing. A summary contract provides hashes for safety [32].

Similar changes were designed in MedicoHealth, MeFy, and MediBloc. MeFy promised to provide physicians with entry access and to protect patients' privacy and anonymity. MediBloc aids the reuse of old medical data history to provide more choice on similar occasions when medical references are needed [33]; they proposed a scheme for drug history queries for physicians, nurses, patients, pharmacists, and management to promise safety and efficiency of the supply chain.

Kuo et al [34] suggested a basic structure in which a Merkle tree prunes the transactions in the tree. The root and header of the tree were designed for easy verification and safety so that an attacker could not undertake the cost of recreating blocks. Ethereum, Hyperledger, and MultiChain are additional emerging applications of blockchain that adhere to standards of the health informatics community. The PoW consensus protocol has been widely adopted. Zcash, Dash, and Monero focus on privacy or anonymity of transaction information for users such as patients.

Many distributed ledger technology (DLT) platforms were examined in [35], such as Bitcoin for digital currency conducted by consensus algorithm; Ethereum virtual machine using smart contracts; Multichain, which uses an application programming interface (API) and command line interface that are easy to interact with to maintain and deploy DLT systems; and EOS, a widely known delegated proof-of-stake cryptocurrency platform with an algorithm consisting of 21 block producers. Cardano supports smart contracts and decentralized applications using a Follow the Satoshi algorithm to introduce a certain amount of randomness without relying on any PoW consensus algorithms. Also, there are many Hyperledger medical blockchain platforms, such as Hyperledger Fabric, which is constructed within the Hyperledger ecosystem with a modular design employing an identity layer and channel layer architecture supported by simplified Byzantine fault tolerance; Hyperledger Sawtooth, which uses a proof-of-elapsed-time algorithm that depends on Intel Software Guard Extensions under a new type of trusted execution environment; and Hyperledger Burrow. Additional platforms use consensus-building and new permissions for verification and security with different data structures, such as Directed Acyclic Graph; these platforms include IOTA, Corda, and Waltonchain. Our research specifically focuses on Ethereum.

System Requirements and System Architecture

The length of the blockchain itself is difficult to decrease; an excessively long chain can influence the transparency efficiency. Lightweight blockchain divides some data from traditional blocks. Lightweight blockchain uses file systems to decrease the initial data storage scale. Notably, because the blocks are decreased, the chain is also cut shorter. The new chain in the lightweight structure is lighter than that in traditional blockchain. After the new file system is introduced, the search efficiency controlled by the traditional data scale can be improved because the chain is shorter. In this paper, we used a lightweight design.

System Requirements

Overall Requirements

In a TCM traceability system, generally speaking, the system should meet the following requirements: realizing the whole traceability process of drugs from planting and processing to distribution to hospital to patients; provide lightweight blockchain to prevent data explosion; ensure the safety of the blockchain to protect the privacy information of all categories of users; and increase the response speed of the blockchain, enabling users to trace authorized information.

In a private chain that provides authorized users with access for verifying blocks, the information security and reduce network pressure must be protected.

User Requirements

In a TCM traceability system based on blockchain, there are four types of users: enterprises (eg, planting enterprise users, production enterprise processing users, sales enterprise users), hospitals, patients, and regulatory departments.

For the enterprise, the system should meet the following requirements:

1. The enterprise should have access to view the basic information of the drugs (excluding privacy) at all stages.
2. The sensitive information of the enterprise should be protected and not be leaked in the system.

For the hospital, the system should meet the following requirements:

1. Users are able to view the drug planting and processing situation.

2. If problems are identified with the drug, the information provided in the traceability system is provided as legal evidence.

For the regulatory authorities, the system should meet the following requirements:

1. When the authorities view all the data in the system, the data are equivalent to the traditional DBA.
2. The authorities can verify that the data in the system are legal and effective.
3. If there is a problem with a drug, the system should provide accountability according to the system information.

For patients, the system should meet the requirement that the patients can view the planting, processing, and sales information of the drugs they purchased.

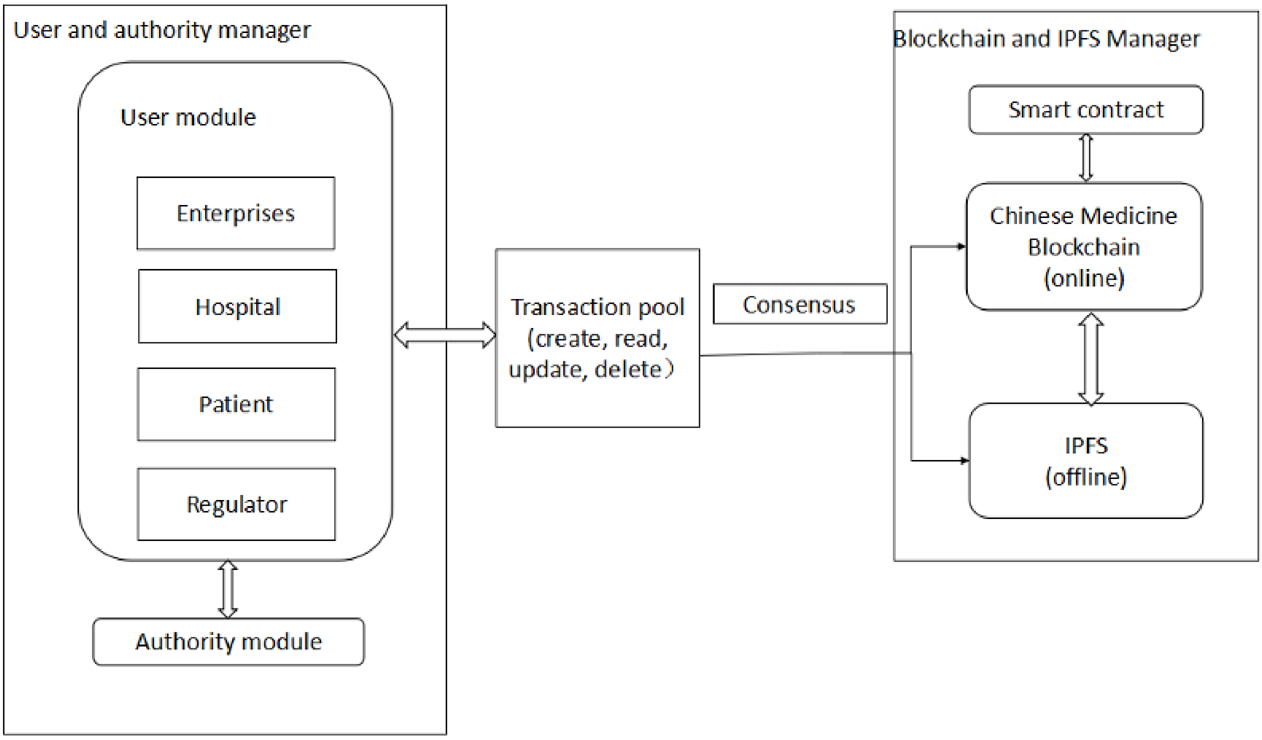
If there is a problem with the drugs, the information provided in the traceability system can be used as legal evidence.

The end users are mainly government staff members controlling the safety and normal transaction of drug trades. The representative government end users are law enforcement personnel, such as SFDA and health commission employees. In China, these data should be administrated by the National Health Commission of the People’s Republic of China.

System Architecture

As shown in Figure 1, the TCM traceability system based on blockchain consists of three components and their relationships:

Figure 1. The architecture of the Chinese medicine traceability system based on blockchain.



1. User module and authority module. The private chain is used to register, manage and authenticate the users who join the blockchain. The authenticated users have different permissions on the system.
2. Transaction tools. The user activities, such as adding, deleting, modifying, and checking, generated in the TCM process are stored in the transaction pool, and these activities will be added to the blockchain and the IPFS.
3. Blockchain and IPFS manager. A certain number of activities in the transaction pool are selected in chronological order. After hashing and packaging, these activities are added to the blockchain and IPFS under the consensus mechanism action. Blockchain mainly stores the Merkle tree and signature of these activities; detailed TCM information is stored in the IPFS. This method avoids information explosion and prevents disclosure of private drug information.

User and Authority Manager

The module consists of two submodules: user and authentication. Users consist of enterprises (planting enterprise, production enterprise, processing enterprise, and circulation enterprise), hospitals, patients, and regulatory departments. Due to the private chain mode, users must be authenticated to participate in the blockchain's activities. The authentication module is equivalent to the certificate authority, which provides effective authentication for users joining the blockchain network. Our system not only stores all types of transactions generated by users, such as information related to the planting enterprise for planting of certain medicinal materials, but also stores the information related to modification and deletion of medicinal materials by users in the transaction pool.

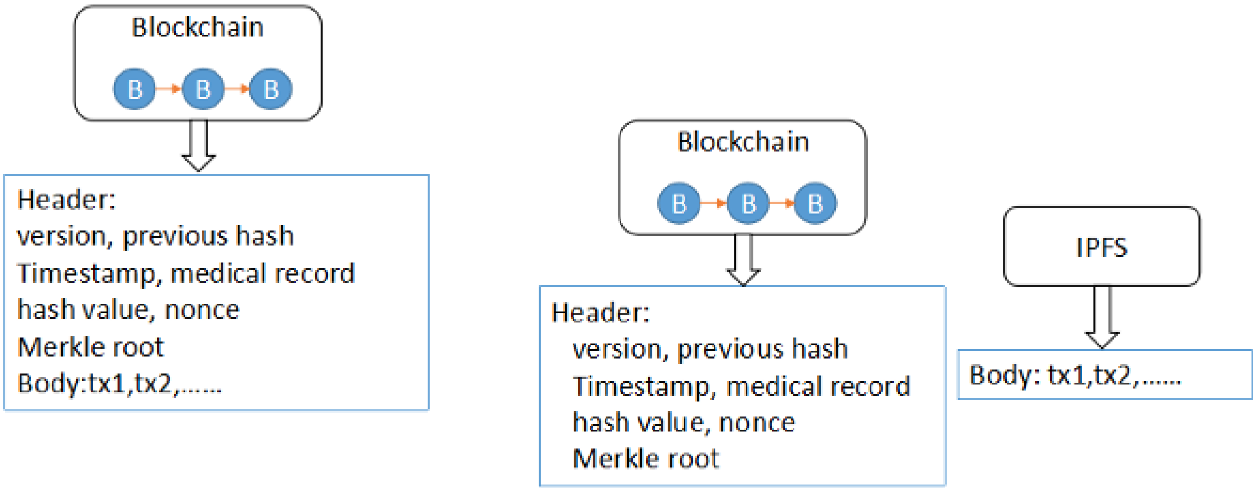
Blockchain and IPFS Management

In Bitcoin, all nodes need to copy the entire blockchain to the local system (all the nodes maintain a replicated copy of the ledger). As the number of blockchain accounts increases, the following problems will arise:

1. Space: the storage space that nodes need to provide increases with the increasing number of account books.
2. Network: each node needs to download the account book, the network cost increases sharply.
3. Calculation: the calculation cost of the nodes in the verification process increases.

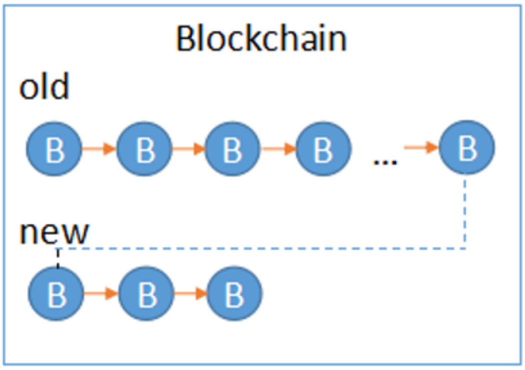
To solve this problem, we have improved the blockchain in two aspects. First, horizontal segmentation of blockchain was adopted. The traditional block, as shown in Figure 2, consists of two parts: a header and a body. The header stores the Merkle tree, timestamp, previous hash, and other parts. The header is mainly used to verify the correctness of the transactions. The body stores the packaged specific transactions; this part occupies most of the space of the blockchain. In the improved blockchain, only the header in the original block is saved, while the body is stored in the IPFS; this design greatly reduces the size of the blockchain.

Figure 2. Blockchain structure. (A) Traditional blockchain structure. (B) Horizontal blockchain structure. IPFS: InterPlanetary File System; tx: text.



Second, vertical segmentation of the blockchain is adopted. The blockchain is formed with blocks according to the time sequence. As the time increases, the length of the blockchain increases correspondingly; this will lead to a sharp increase in the storage and calculation costs of the nodes. The improved blockchain is divided into two segments, as shown in Figure 3.

Figure 3. Vertical blockchain structure.



The new blockchain only stores the blocks added recently, and the old segment stores the past blocks. New and old blocks are connected by pointers. When nodes verify the newly added blocks, there is no need to copy all the blockchains; instead, only the new segment blocks are copied.

This approach greatly improves the verification efficiency and system performance. New and old blocks can be designed to be segmented according to time or size. If time is rotated as the threshold, the blocks are segmented by month, quarter, or even year according to the actual situation. Once the time node is reached, the old blocks generated are stored in the IPFS, and the new blocks are added to the new blockchain.

Transaction Pool

This module has two functions: first, the information is generated by various users (eg, enterprises, hospitals, and patients) in the user module. Second, the miner selects a certain amount of transactions from the transaction pool and adds them to the blockchain.

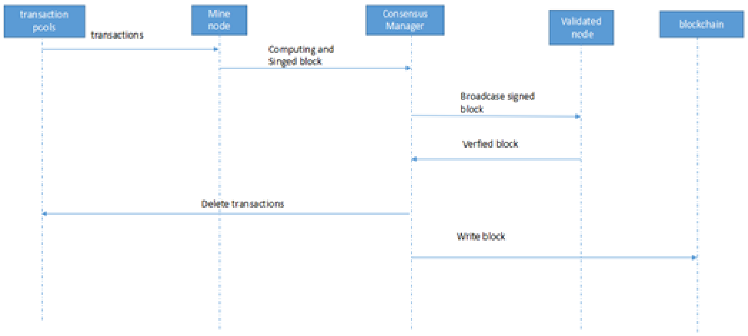
Once these transactions are verified by the consensus mechanism, they are added to the blockchain and the selected transactions are deleted.

System Implementation

The TCM data were collected from participants in the drug trade, including farmers, companies, and drugstores; two popular representative drugs were chosen for testing. The data contained each step in the drug production process. The data were generated from these parts: drug planting at the base, growing, harvesting, processing, packaging, storage, circulation, and selling to patients. The data were generated from these parties: TCM dealers, patent TCM pharmaceutical production companies, and raw material sources. The TCM dealer data consisted of the price, approval number, shop name, province, patent TCM pharmaceutical production medicine name, TCM medicine under quality control, and date of sale. The medicine enterprise data consisted of the names of the medicinal products, main raw materials, production enterprises, regions, and TCM medicine under quality control. The raw material source data included the names of the production companies, province, scope of business, and herbal TCM medicine under quality control. A total of thousands of pieces of data based on two types of herbal TCM medicine in circulation were used in our experiments.

Figure 4 shows how the activities in the transaction pool are added to the blockchain and IPFS. Blockchain nodes are divided into mining nodes and verification nodes. All users in the blockchain are used as mining nodes or verification nodes:

Figure 4. Data flow of the blockchain and IPFS manager. IPFS: InterPlanetary File System.



1. Mining node (equivalent to miners): after mining starts, the transactions in the transaction pools (also stored in the database) are organized to form a pending block; this process is a hash calculation that conforms to the nonce rule. Once the calculation is successful, the block is submitted to consensus management and the blocks submitted by other miners are rejected by the consensus. Then, consensus management starts consensus processing and publishes an announcement to enable verification of the verification node. If the block is successfully verified, it is added to the blockchain.
2. Verifying the nodes. After receiving the announcement issued by consensus management, the pending transaction block is verified (this verification is very fast) and signed, and it is then sent to consensus management. According to consensus management statistics, if the threshold value (2/3) for joining the blockchain is reached, the header and body of the block are added to the blockchain and IPFS respectively, and the transactions that have been added to the block are deleted from the transaction pool. If this process is successful, the block joins the blockchain.

Our system, the smart contract empowered system, is implemented on the open source Ethereum platform. This platform includes five parts: the blockchain, IPFS (version 0.4.13), smart contract, Ethereum decentralized application (DAPP) platform and the third-party MetaMask wallet. The Ethereum DAPP application platform is composed of Node.js (version 8.9.4) and Truffle (version 5.1.12). It is implemented with the Truffle project framework, mainly through calling the relevant API to connect each part and import the relevant blockchain account information, compile and deploy the smart contract, and complete the relevant user requests. The smart contract code is written in solid language, and the Truffle contract API is used to synchronize transactions and deploy to the chain. The main function of a smart contract is to code according to the business rules of different identity registrations, traceability information upload, authorization and revocation of related roles, and query and access of end users, as well as to manage the corresponding permissions. The IPFS is mainly responsible for storing the encrypted privacy information of the Chinese herbal medicine supply chain. The IPFS performs synchronization of the data of the nodes and returns the final hash routing address. Finally, the blockchain is responsible for storing the smart contract code, completing the call of the relevant smart contract interface, packaging the execution result of the request in the form of the blocks, and writing the transaction records into the ledger after the node consensus verification.

A smart contract in Ethereum is deployed by means of a transaction by spending a certain amount of ether, the “cryptocurrency” of the Ethereum network, and invoked with input data via a transaction. In our system, the transaction is related to a TCM drug. The smart contract code is written in solidity language (version 0.4.21).

Details of the algorithms of traceability information data upload, authority control of traceability, and inquiry and access of traceability are provided in Textboxes 1-3, respectively.

Textbox 1. Algorithm 1: traceability information data upload.

Input: TCM herbal medicine information, Blockchain length threshold L

Output: The hash returned by the InterPlanetary File System (IPFS), the block upload success information

Acquisition phase

1. A unique identification is assigned for a single traditional Chinese medicine (TCM) herbal medicine or batch by radio frequency identification tag, and the labels are sent to manufacturer A’s database.

User registration

2. The node initiates the registration information request.
3. The blockchain account address information and public private key pair are generated.
4. The smart contract-registration interface is called.
5. The enterprise deposits the initial transaction amount to the blockchain account address.

Upload request

6. The nodes input the TCM transaction.
7. The private key signs the request, and the encrypted information is stored in the IPFS.
8. The corresponding hash is returned.

Blockchain nodes verification

9. The traceability information attribute and encrypted information hash value are passed in as parameters.
10. The smart contract is executed and broadcasts the unique identification code of the bound enterprise, encrypted information hash value, and traceability information attribute.

Calculation of the length of the blockchain

11. The blockchain is divided into old and new parts, and the contract execution results are written to the new blocks.

Return

12. The blockchain upload success information is returned.

Textbox 2. Algorithm 2: authority control of traceability.

Input: Traceability user and authorization information attributes

Output: Success tips

Authorization:

1. The user inputs the corresponding traceability user and authorization information attributes.
2. The user initiates the authorization request and encrypts the request with the private key.
3. The Ethereum application platform requests the blockchain nodes to call the smart contract authorization interface.
4. The request is broadcast to other nodes synchronously and the authorization information is written into the account book records.
5. The Ethereum application platform uses a public key to encrypt the corresponding key.
6. The user decrypts the key through the private key.

Textbox 3. Algorithm 3: inquiry and access of traceability.

Input: The unique identification ID, traditional Chinese medicine (TCM) medicinal material identification ID and traceability information attribute of the traceability information owner enterprise

Output: The hash value of successful prompt corresponding to the query information, encrypted privacy traceability information

Results

1. The user inputs the unique identification ID, TCM herbal medicine identification ID, and traceability information attribute of the owner of the traceability information.

2. Ethereum extracts information attributes from the request interface and executes the contract.

3. Ethereum sends a request query to the InterPlanetary File System node and returns the encrypted privacy traceability information.

Return

4. The results are returned.

Performance Evaluation

The construction of the TCM traceability system is mainly based on the Ethereum platform to build a public blockchain and private chain blockchain. Four Ubuntu cloud servers with the Geth client were built as different account addresses using Go language; the creation block information and Geth configuration files of the four nodes were configured, and Ganache-cli was used to initialize accounts and generate public and private key pairs. Four server nodes were formed into blockchain network nodes, and the underlying consensus trading environment was achieved. The Ethereum platform was built with the Truffle framework, and the Remix integrated development environment (IDE) visualization tool was used to perform the corresponding account address entry, compilation and deployment of the smart contract, and method call of the smart contract interface. IPFS file storage was mainly achieved by writing Web3.js script in the Truffle framework and calling the IPFS API. The core file of the whole framework was mainly Web3.js script. The core contained a smart contract address, IPFS calling interface, related function methods, and page code. The Geth + Truffle + IPFS terminal combination constituted the software environment of the traceability system. The hardware environment mainly consisted of a notebook computer, Core i7 processor (Intel Corporation), 16 GB memory, and 1 TB hard disk space. Three parts of the time and space capabilities of the system were tested: (1) the relationship between the storage space of the blockchain traceability system, IPFS, and new TCM traceability system with the upload times; (2) the relationship of the changes in the upload response time with the block capacity; and (3) the relationship between the changes in the query response time and the block capacity.

The results of these three time and space capacity tests showed that the processing level of the new TCM traceability system increases in time and space. The results were analyzed to further verify the comprehensive ability of the system.

We uploaded the same traceable drug file multiple times in three different systems. Figure 5 reveals that although the common blockchain system has consensus traceability ability, the data on the chain will continue to increase steadily; this will eventually lead to performance problems and data security problems due to a long chain and data volume oversaturation. The storage space of the IPFS system increases steadily. The generalized Merkle directed acyclic graph structure endows the system with unlimited storage space. The IPFS structure will partition the block according to the file size and finally return the hash address. The new TCM traceability system combines the advantages of the common system and IPFS. The new TCM system greatly reduces the data storage capacity on the chain. When the space resources are optimally used, security problems caused by the data explosion on the chain are also prevented. Storing the business data on our system’s chain in a third-party distributed database further improves its information traceability and privacy protection abilities.

Figure 5. Comparison of the increase of the storage space according to the size of the data upload.

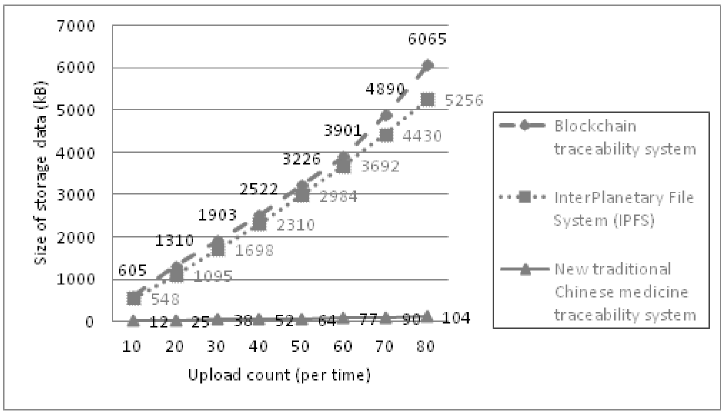


Figure 6 reveals that the IPFS system has a strong upload response capability. The ability of IPFS to deal with bandwidth problems caused by the block capacity increase is also relatively obvious. The common blockchain traceability system obviously affects the response speed of the upload requests due to the continuous growth of the block capacity, and the response time will increase exponentially with the rapid growth of the block capacity. This increase creates response delay and affects the transaction writing process. Therefore, the new TCM traceability system improves the upload response speed of the system by combining it with IPFS technology and makes full use of the data distribution ability of hash tables to improve the user traceability effects. Due to the impact of the bandwidth performance and the complexity of the business logic code calling interface, the upload time is slow, so the efficiency does not reach the maximum; however, the overall optimization trend is obvious.

Figure 6. Comparison of the upload response times with different sizes of block storage.

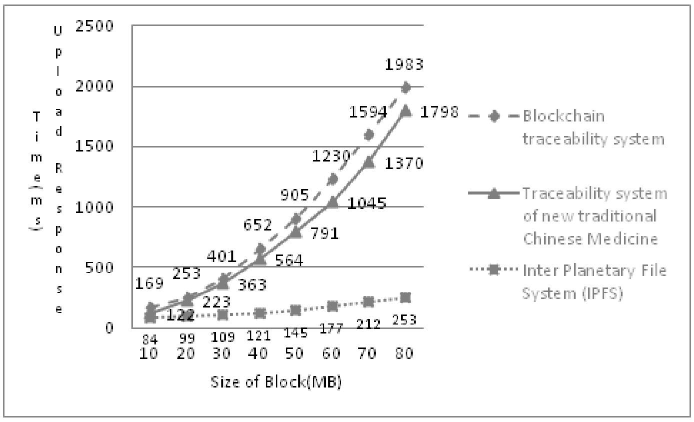
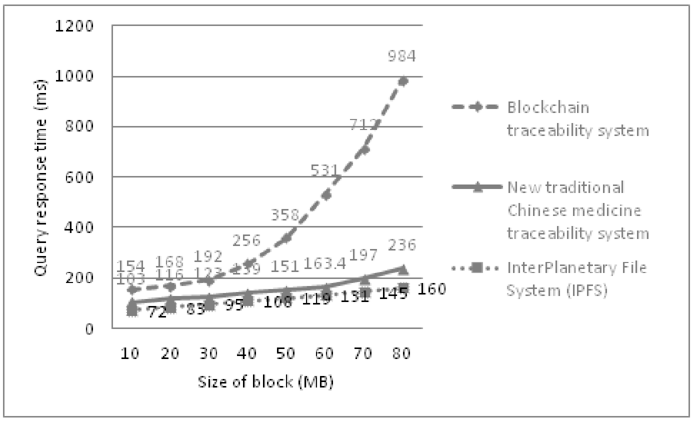


Figure 7 reveals that the query response time of the general blockchain traceability system is greatly affected by the block capacity. As the newly added transaction data are continuously written into the blockchain ledger, the number of blocks on the chain increases. When the query traceability information interface is called, the blockchain will search for data on the chain step by step; however, if the data have been hashed to form a Merkle tree structure, the response time becomes slower. The query ability of the IPFS system is very strong and stable. The new TCM traceability system combined with IPFS technology can greatly improve the response performance. Because the system only needs to store the hash address returned by the IPFS network, it controls the amount of business data on the chain to the minimum level, thus reducing the recursive query time of the system and obtaining the Merkle tree hash value; therefore, users enjoy the simplest and fastest information traceability experience.

Figure 7. Comparison of the query response time changes with different sizes of block storage.



Discussion

Blockchain is a type of tamperproof distributed accounting technology. Blockchain possesses natural advantages when applied to the traceability of TCM herbal medicine. However, compared with the existing Bitcoin and Ethereum blocks, there is a much larger amount of information related to TCM herbal medicine, leading to a much higher risk of data explosion. In addition, some privacy information regarding TCM herbal medicine must be protected. In this paper, we designed a decentralized system based on blockchain and IPFS. Horizontal and vertical segmentation are conducted in the blocks to greatly decrease the data in our system; these segmentations solve the data explosion problem and protect privacy information.

The users of related drug administration input the unique identification ID, TCM herbal medicine identification ID, and traceability information attribute of the owner of the traceability information for the query system. Then, Ethereum extracts the information attributes from the request interface and executes the contract. After Ethereum sends the request query to the IPFS node and returns encrypted privacy traceability information, the results are returned. Administrative organizations such as the SFDA can obtain all the information they need to achieve chain data traceability for further investigation.

An experiment was conducted to verify the efficiency of the new method, and our method was also compared with other methods (Table 1).

Table 1. Performance comparison chart of the new TCM traceability system and current mainstream traceability systems.

Characteristic	Centralized traceability system	EPCIS[a] network service traceability system	Blockchain traceability system	New TCM[b] traceability system
Information traceability ability	Low	High	High	High
Tamperproof capability	Low	Low	High	High
Privacy protection capability	Low	Low	Low	High
Data dispersion capability	Low	Low	Low	High
Chain data size	N/A[c]	N/A	Large	Small
Data response speed	N/A	N/A	Medium	Fast

[a] EPCIS: Electronic Product Code Information Services.

[b] TCM: traditional Chinese medicine.

[c] N/A: not applicable (incomparable).

The data were collected from aspects of the drug trade for computer validation. Compared with Ethereum, our experimental results demonstrated that our proposed architecture shows superior performance in the following aspects:

1. The speed of uploading data to the blockchain is faster.
2. The speed of querying the blockchain improves.
3. Network traffic is improved because users only need to copy part of the blockchain.

In the future, we would extend the proposed work in the following aspects:

1. User node optimization: in this paper, transaction information is stored in IPFS, which may affect the system performance to a certain extent. In future work, we would add some super user nodes to store not only old blocks but also some transaction information.
2. Blockchain query optimization: as the blockchain length increases, the query efficiency decreases sharply. Therefore, a structure similar to an index would be designed in future work to accelerate the query performance of the blockchain.
3. Privacy information protection: data related to drugs contain much privacy information, which the method proposed in this paper does not consider. Therefore, we expect to implement an attribute-based signature encryption mechanism in the smart contract to further protect privacy information.

Acknowledgments

Previous medical blockchain applications combined the advantages of blockchain with applications in business, agriculture, and food; these studies are the basis of our research. We are thankful to all the researchers and applications in the blockchain field as well as for the work in other fields that led us to develop a TCM herbal medicine system. This work is partially supported by the Opening Project of Guangdong Province Key Laboratory of Big Data Analysis and Processing at Sun Yat-sen University (202001) and the Social Sciences Project of Guangzhou University of Chinese Medicine (grants 2020SKYB05 and 2020SKXK25).

Conflicts of Interest

None declared.

References

1. Notice of the office of the State Administration of Traditional Chinese Medicine for national traditional Chinese medicine standardization project of application (EB/OL). Webpage in Chinese. State Administration of Traditional Chinese Medicine. 2015 Jul 17. URL: http://kjs.satcm.gov.cn/gongzuodongtai/2018-03-24%20/3257.html
2. Shi MY, Wen CB, Zhao ST. Analysis on the present situation of traceability system of traditional Chinese medicine. Journal of Chengdu University of Traditional Chinese Medicine 2016:04.
3. Announcement of the State Food and Drug Administration on doing a good job in the construction of the informatization traceability system for key varieties (No. 111, 2020). Webpage in Chinese. National Medical Products Administration. 2020. URL: https://www.nmpa.gov.cn/xxgk/ggtg/qtggtg/20201013155906186.html [accessed 2021-06-11]
4. Pilot Project Program Under the Drug Supply Chain Security Act; Request for Comments. US Food and Drug Administration. URL: https://s3.amazonaws.com/public-inspection.federalregister.gov/2017-15203.pdf [accessed 2021-06-11]

5. Ding Jinxi J, Lu H, Meng LL. European Union Drug Electronic Supervision System and its enlightenment. Article in Chinese. Chinese Journal of Pharmaceuticals 2012;43(8):718-A66 [FREE Full text]
6. Notice of the office of the State Administration of Traditional Chinese Medicine for national traditional Chinese medicine standardization project of application (EB/OL). Webpage in Chinese. State Administration of Traditional Chinese Medicine. 2015 Jul 17. URL: http://kjs.satcm.gov.cn/gongzuodongtai/2018-03-24%20/3257.html [accessed 2021-06-07]
7. Guiding opinions of SFDA on the construction of drug information traceability system. Chinese State Food and Drug Administration. URL: http://www.gov.cn/zhengce/zhengceku/2018-12/31/content_5434073.htm [accessed 2021-06-10]
8. Announcement of Chinese State Food and Drug Administration for issues of 3 kinds of informationize standard including basic technical requirements of drug traceability system. Chinese State Food and Drug Administration. 2019 Aug 23. URL: https://www.nmpa.gov.cn/yaopin/ypggtg/ypqtgg/20190827171501407.html [accessed 2021-06-07]
9. Angeles R. Blockchain-based healthcare: three successful proof-of-concept pilots worth considering. J Inf Technol Manag 2019;27(3) [FREE Full text]
10. Tian F. 2016 Presented at: 2016 13th International Conference on Service Systems and Service Management (ICSSSM); June 24-26, 2016; Kunming, China. [doi: 10.1109/icsssm.2016.7538424]
11. Galvez JF, Mejuto J, Simal-Gandara J. Future challenges on the use of blockchain for food traceability analysis. Trends Analyt Chem 2018 Oct;107:222-232. [doi: 10.1016/j.trac.2018.08.011]
12. Toyoda K, Mathiopoulos PT, Sasase I, Ohtsuki T. A novel blockchain-based product ownership management system (POMS) for anti-counterfeits in the post supply chain. IEEE Access 2017;5:17465-17477. [doi: 10.1109/access.2017.2720760]
13. Aitken R. Forbes. 2017. URL: https://www.forbes.com/sites/rogeraitken/2017/12/14/ibm-walmart-launching-blockchain-food-safety-alliance-in-china-with-fortune-500s-jd-com/#556eb5fe7d9c [accessed 2021-06-07]
14. Tse D, Zhang B, Yang Y, Cheng C, Mu H. Blockchain application in food supply information security. 2017 Presented at: 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); December 10-13, 2017; Singapore. [doi: 10.1109/ieem.2017.8290114]
15. San Miguel E, Timmerman R, Mosquera S, Dimogerontakis E, Freitag F, Navarro L. Blockchain-enabled participatory incentives for crowdsourced mesh networks. In: Djemame K, Altmann J, Bañares J, Agmon Ben-Yehuda O, Naldi M, editors. Economics of Grids, Clouds, Systems, and Services. GECON 2019. Lecture Notes in Computer Science, vol 11819. Cham, Switzerland: Springer; Nov 24, 2019.
16. Mamo N, Martin GM, Desira M, Ellul B, Ebejer J. Dwarna: a blockchain solution for dynamic consent in biobanking. Eur J Hum Genet 2020 May;28(5):609-626 [FREE Full text] [doi: 10.1038/s41431-019-0560-9] [Medline: 31844175]
17. Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and trustable electronic medical records sharing using blockchain. AMIA Annu Symp Proc 2017;2017:650-659 [FREE Full text] [Medline: 29854130]
18. Olson LE, Rosulek MJ, Winslett M. A generalized honest-but-curious trust negotiation strategy for harvesting credentials. Illinois Digital Environment for Access to Learning and Scholarship. 2009 Apr 22. URL: http://hdl.handle.net/2142/11381 [accessed 2021-06-11]
19. Hang, Choi, Kim. A novel EMR integrity management based on a medical blockchain platform in hospital. Electronics 2019 Apr 25;8(4):467. [doi: 10.3390/electronics8040467]
20. Krittanawong C, Rogers AJ, Aydar M, Choi E, Johnson KW, Wang Z, et al. Integrating blockchain technology with artificial intelligence for cardiovascular medicine. Nat Rev Cardiol 2020 Jan 11;17(1):1-3. [doi: 10.1038/s41569-019-0294-y] [Medline: 31605093]
21. Wu R, Ahn G, Hu H. Secure sharing of electronic health records in clouds. 2012 Presented at: 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom); October 14-17, 2012; Pittsburgh, PA. [doi: 10.4108/icst.collaboratecom.2012.250497]
22. Nguyen DC, Pathirana PN, Ding M, Seneviratne A. Blockchain for secure EHRs sharing of mobile cloud based e-health systems. IEEE Access 2019;7:66792-66806. [doi: 10.1109/access.2019.2917555]
23. Ismail L, Materwala H, Zeadally S. Lightweight blockchain for healthcare. IEEE Access 2019;7:149935-149951. [doi: 10.1109/access.2019.2947613]
24. Lewandowski M, Płaczek B. An event-aware cluster-head rotation algorithm for extending lifetime of wireless sensor network with smart nodes. Sensors (Basel) 2019 Sep 20;19(19):4060 [FREE Full text] [doi: 10.3390/s19194060] [Medline: 31547047]
25. Arjona R, Prada-Delgado M, Arcenegui J, Baturone I. A PUF- and biometric-based lightweight hardware solution to increase security at sensor nodes. Sensors (Basel) 2018 Jul 26;18(8):2429 [FREE Full text] [doi: 10.3390/s18082429] [Medline: 30049967]
26. Lwin MT, Yim J, Ko Y. Blockchain-based lightweight trust management in mobile ad-hoc networks. Sensors (Basel) 2020 Jan 27;20(3):698 [FREE Full text] [doi: 10.3390/s20030698] [Medline: 32012774]
27. Fu J, Wang N, Cai Y. Privacy-preserving in healthcare blockchain systems based on lightweight message sharing. Sensors (Basel) 2020 Mar 29;20(7):1898 [FREE Full text] [doi: 10.3390/s20071898] [Medline: 32235389]

28. Rangachari P, Dellsperger KC, Rethemeyer RK. A qualitative study of interprofessional learning related to electronic health record (EHR) medication reconciliation within a social knowledge networking (SKN) system. JHL 2019 Mar;Volume 11:23-41. [doi: 10.2147/jhl.s198951]
29. MedicoHealth Whitepaper. MedicoHealth. 2018. URL: https://medicohealth.io/supporters/documents/wp_beta.pdf [accessed 2019-03-15]
30. MeFy Whitepaper. MeFy. 2018. URL: https://gust.com/companies/mefyio [accessed 2019-03-15]
31. MediBloc Technical Whitepaper. GitHub. 2018. URL: https://medibloc.org/en [accessed 2021-06-07]
32. Azaria A, Ekblaw A, Vieira T, Lippman A. Medrec: using blockchain for medical data access and permission management. In: Proceedings of the 2nd International Conference on Open and Big Data (OBD). 2016 Presented at: 2nd International Conference on Open and Big Data; August 22-24, 2016; Vienna, Austria p. 25-30. [doi: 10.1109/obd.2016.11]
33. Jamil F, Hang L, Kim K, Kim D. A novel medical blockchain model for drug supply chain integrity management in a smart hospital. Electronics 2019 May 07;8(5):505. [doi: 10.3390/electronics8050505]
34. Kuo T, Zavaleta Rojas H, Ohno-Machado L. Comparison of blockchain platforms: a systematic review and healthcare examples. J Am Med Inform Assoc 2019 May 01;26(5):462-478 [FREE Full text] [doi: 10.1093/jamia/ocy185] [Medline: 30907419]
35. Chowdhury MJM, Ferdous MS, Biswas K, Chowdhury N, Kayes ASM, Alazab M, et al. A comparative analysis of distributed ledger technology platforms. IEEE Access 2019;7:167930-167943. [doi: 10.1109/access.2019.2953729]

Abbreviations

API: application programming interface
DAPP: decentralized application
DL: distributed ledger
DLT: distributed ledger technology
DoS: denial of service
EHR: electronic health record
EMR: electronic medical record
IDE: integrated development environment
IPFS: InterPlanetary File System
PoW: proof of work
SFDA: State Food and Drug Administration
RFID: radio-frequency identification
RSA: Rivest-Shamir-Adleman
TCM: traditional Chinese medicine

Edited by R Kukafka; submitted 22.11.20; peer-reviewed by Z Su, TT Kuo; comments to author 19.12.20; revised version received 10.02.21; accepted 04.05.21; published 21.06.21.

Please cite as:
Wang Z, Wang L, Xiao F, Chen Q, Lu L, Hong J
A Traditional Chinese Medicine Traceability System Based on Lightweight Blockchain
J Med Internet Res 2021;23(6):e25946
URL: https://www.jmir.org/2021/6/e25946
doi:10.2196/25946
PMID:34152279

Viewpoint

Why Are Digital Health Care Systems Still Poorly Designed, and Why Is Health Care Practice Not Asking for More? Three Paths Toward a Sustainable Digital Work Environment

Johanna Persson[1*], MSc, PhD; Christofer Rydenfält[1*], MSc, PhD

Department of Design Sciences, Lund University, Lund, Sweden

[*]all authors contributed equally

Corresponding Author:
Johanna Persson, MSc, PhD
Department of Design Sciences
Lund University
Box 118
Lund, 22100
Sweden
Phone: 46 462224358
Email: johanna.persson@design.lth.se

Abstract

Knowledge of how to design digital systems that are ergonomically sound, high in usability, and optimized for the user, context, and task has existed for some time. Despite this, there are still too many examples of new digital health care systems that are poorly designed and that could negatively affect both the work environment of health care staff and patient safety. This could be because of a gap between the theoretical knowledge of design and ergonomics and the practical implementation of this knowledge in procuring and developing digital health care systems. Furthermore, discussions of digitalization are often at a general level and risk neglecting the nature of direct interaction with the digital system. This is problematic since it is at this detailed level that work environment and patient safety issues materialize in practice. In this paper, we illustrate such issues with two scenarios concerned with contemporary electronic health care records, based on field studies in two health care settings. We argue that current methods and tools for designing and evaluating digital systems in health care must cater both to the holistic level and to the details of interaction and ergonomics. It must also be acknowledged that health care professionals are neither designers nor engineers, so expectations of them during the development of digital systems must be realistic. We suggest three paths toward a more sustainable digital work environment in health care: (1) better *tools for evaluating the digital work environment in the field*; (2) *generic formulations of qualitative requirements* related to usability and for adaptation to the user, context, and task, to be used in procurement; and (3) the introduction of *digital ergonomics* as an embracing concept capturing several of the ergonomic challenges (including physical, cognitive, and organizational aspects) involved in implementing and using digital systems.

(J Med Internet Res 2021;23(6):e26694) doi:10.2196/26694

KEYWORDS

digital systems; electronic health records; digital work environment; ergonomics; usability; human-centered design

Introduction

There is no doubt that the digital transformation of health care is changing how health care is delivered [1,2]. Besides aspects of efficiency, safety, and patient empowerment, digitalization also affects the work environment of health care staff. New digital systems can both facilitate and constrain work, depending on their design and how well they support organizational goals and objectives [3-5]. Furthermore, technology is never neutral. Verbeek [6] notes the following:

> *When technologies are used, they always help to shape the context in which they fulfill their function. They help to shape human actions and perceptions, and create new practices and ways of living.*

Researchers have studied the impact of digital systems on the work environment since the early eighties [7-11]. Since then, diverse digital systems have emerged, ranging from personal computers for rudimentary office tasks [12] to more sophisticated systems, such as systems for computer-supported collaborative work [13,14], or mobile devices and integrated software and hardware [15,16].

In parallel with the introduction of digital systems in the workplace, the field of human-computer interaction (HCI) has developed, and the usability (ie, the efficient, effective, and satisfactory use of a system with respect to the specified user, goal, and use context [17]) of digital systems in health care has become a broadly used concept [18-26]. The idea of involving users in software development to ensure the creation of systems with relevant functionality and high usability had already been introduced in the eighties [27]. Along with this, HCI design principles emerged, intended to match the system to the user's cognitive abilities, such as limitations of memory, the perceptual system, and mental models [28-30]. A user-centered design process and HCI design principles still constitute the foundation of both HCI education and practice. Nowadays, there is even a standardized process for developing new digital systems intended to ensure that human perspectives are considered [31]. Furthermore, many engineers and developers with HCI knowledge are graduating from higher education institutions worldwide [32-34]. Together, this implies the existence of the technical proficiency and maturity required to develop digital systems with relevant functionality and good usability.

Despite this, basic usability-related problems continue to emerge in health care practice [5,35-37]. Issues related to efficiency, safety, and work environment follow in the tracks of the digital transformation of health care. The problems have even been referred to as "the other health care crisis" (the first one being the COVID-19 pandemic) [38], and links to job frustration and burnout have been investigated [39]. Solutions are hence asked for, and the search for explanations is ongoing [40-42]. We ask ourselves why this is still so when the conditions for developing usable and efficient digital systems, based on HCI design principles and user-centered design methods, have never been better.

In this paper, we identify new ways to overcome obstacles related to poorly designed digital health care systems. Specifically, we highlight the potential effects on the work environment of violations of HCI design principles and of deficient knowledge of the user, use context, and task. This is illustrated through the analysis of two scenarios, providing insights into how these shortcomings materialize in health care practice. Based on insights from these scenarios, we argue for three paths forward that we believe could lead to a more ergonomically sound digitalization of health care.

Effects of Poorly Designed Digital Systems on Health Care Professionals' Work Environment

Two Scenarios

In the following section, we first present two scenarios concerning the use of contemporary electronic health care records (EHRs). EHRs are used as a means to demonstrate our viewpoints, as they are used by basically everyone in the health care sector, whether you are in home care, in primary care, or in specialized care, and whether you are a doctor, a nurse, a lab assistant, or an administrator. They are, furthermore, frequently used throughout the day. We then outline four insights from the scenarios built on basic HCI theory, which illustrate how digital systems affect health care professionals' work environment.

The two scenarios are synthesized from real situations witnessed in data from in vivo observations of Swedish home health care and primary care performed in 2014 and 2018 as part of two research projects. One project aimed at studying digital systems used by practitioners in home health care. The other one studied the implementation of a new EHR for primary care and included both an expert evaluation of the usability of the system as well as in vivo observations of the system at primary care centers.

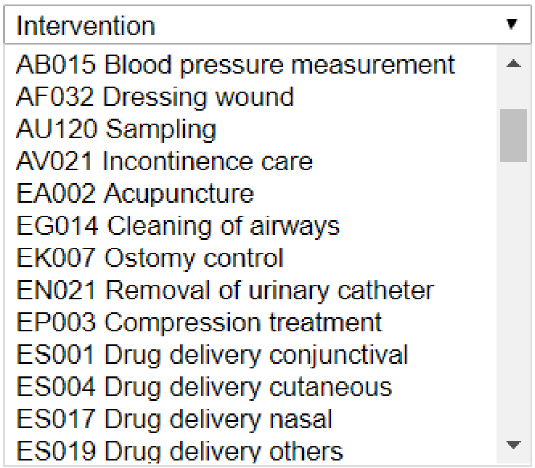
Scenario 1: Home Health Care

On Monday morning, Anna, a 47-year-old registered home health care nurse, logs in to her computer and opens the EHR to get an update on what has happened with her patients since she last saw them on Thursday the week before. She starts by looking through the messages sent by the staff working during the weekend. There are 51 new messages in the message box. With the introduction of the new EHR system, the number of messages has multiplied due to a new way of sharing information in which messages are sent to work groups instead of individuals. Many of the messages are obsolete because they were already addressed during the weekend, while others are not relevant to Anna's work. She understands that she must read 20%-25% of the messages more carefully and that the only way to find them is to go through the list one message at a time and delete the irrelevant ones.

Anna needs to check what has happened to one of her patients during the weekend, so she opens the health record for that patient. The information is sorted chronologically with the oldest post first, and since the patient has had home health care for 3 years, the health record is long. To find the most recent post from the past weekend, Anna must scroll through several pages of records. To avoid this time-consuming activity, which entails a lot of clicking, Anna has learned that she can access the information in reverse chronological order by creating a PDF file for printing. By using this workaround, she can access the information she wants about the patient more quickly.

In the afternoon, Anna returns to the office from a patient visit and opens the EHR to document the visit. This procedure is by far the one most often performed in the system. The process is cumbersome. Besides the actual writing, it takes Anna 19 clicks to find the patient, open the patient's health record, locate the correct place to insert information, and then verify and save the information. To make things even more annoying, the system takes more than a second to respond to each of Anna's clicks, with the consequence that she makes an extra click or two just to be sure. As part of the documentation process, Anna must also classify the intervention performed at the patient's home. Each intervention has a name, for example, "Ostomy control," and a code, for example, "EK007," selected from a drop-down menu with a long list of all available interventions (Figure 1). However, the list of interventions is not sorted by intervention name but by intervention code, which is not in line with how Anna stores the information in her head. Anna starts reading every line in the long list, and when she comes to the end of the page, she scrolls down. Finally, she finds what she is looking for: “EK007 Ostomy control.” Anna marks the list entry using the cursor and clicks “Save.”

Figure 1. An illustration of a drop-down menu with interventions sorted by code and not by name. The user must scroll through the list to find the desired intervention.

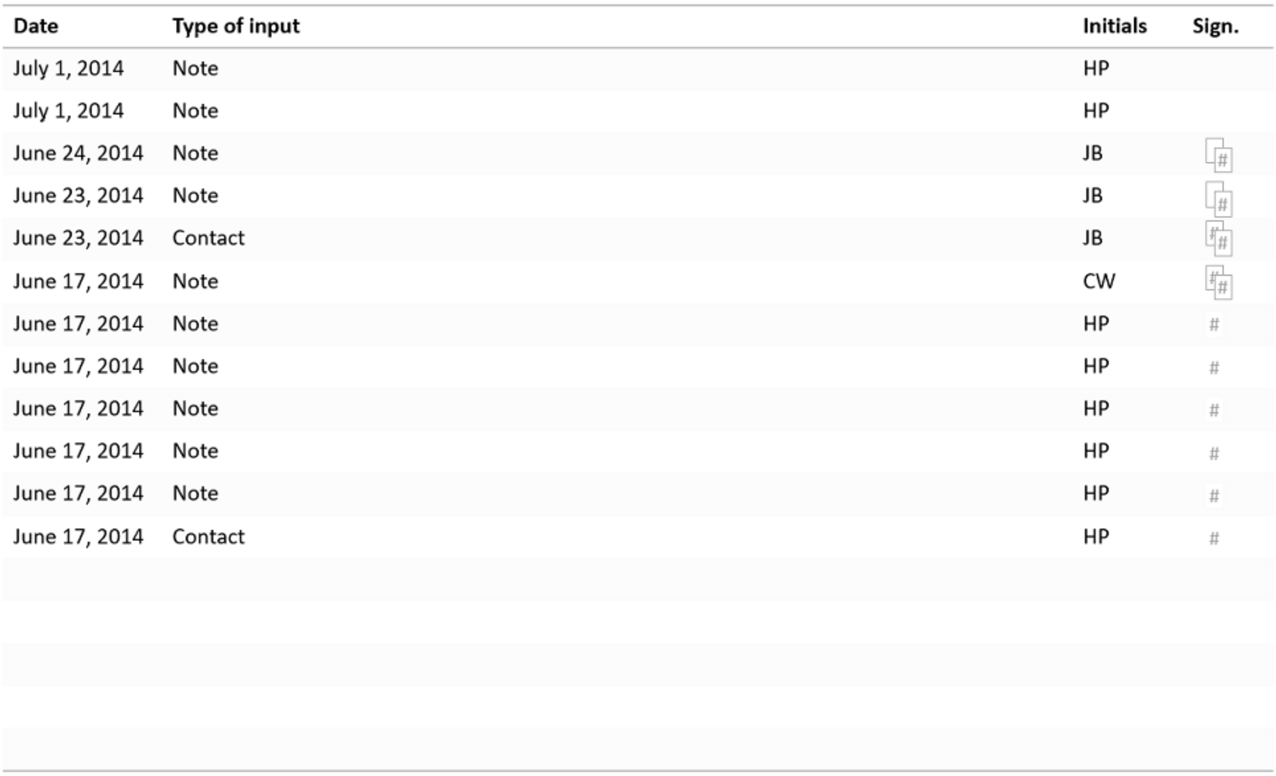


Scenario 2: Primary Care

General practitioner Helen is finishing her shift at the primary care center. Before she can go home, she has a long list of EHR posts to verify and sign. She has postponed this activity for too long since it is so cumbersome to perform. There are several ways this signing can be done, but she has not found a way that suits the workflow and that she can easily remember, so she is unsure whether she is doing it correctly. She needs to stay focused when jumping between the view in which she keeps track of which posts are still unsigned and the view in which the actual information to be signed off is presented. Furthermore, the feedback on each post is difficult to grasp. A post with a certain hashtag symbol has been signed and needs no further attention. Two smaller, overlapping hashtag symbols indicate a post that was signed but has been changed and thus must be viewed and signed again. A post that is unsigned has no symbol at all and is thus not highlighted in any way (Figure 2). The symbols are quite small and appear in different locations in different views in the system (Figure 2, Figure 3), so Helen must search to find them when switching between views.

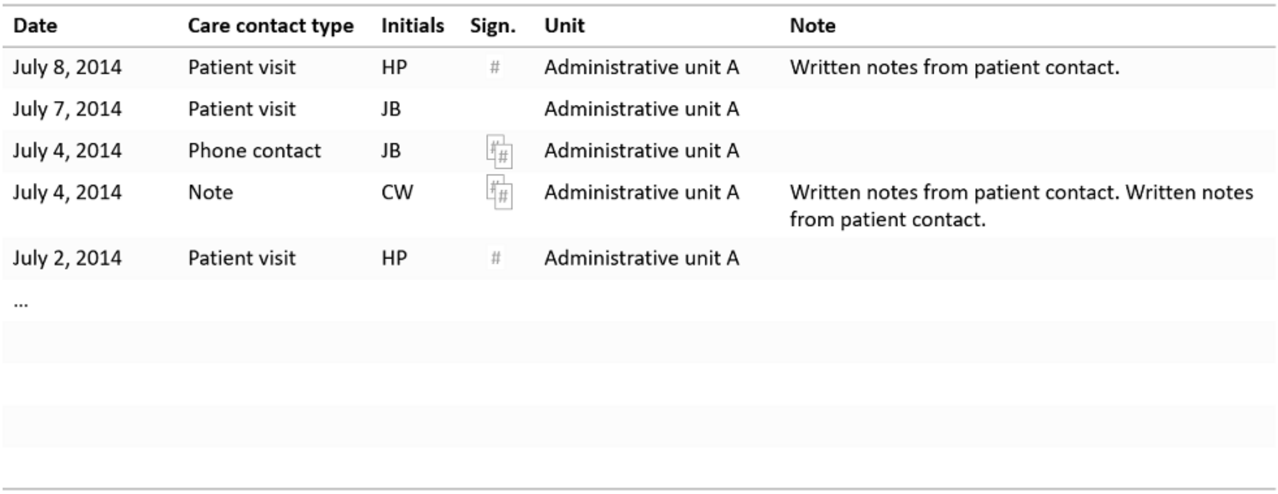
Helen is interrupted in her work by a medical laboratory assistant. He wants Helen to change the lab test order she placed for a patient. The lab test that Helen selected from the list in the EHR system cannot be performed at this primary care center. It is not that she did not know this, but that the list from which lab tests are chosen is long and the names are often similar, so she sometimes clicks on the wrong one. Helen is not the only physician making this mistake, and to prevent this from happening, the medical laboratory has printed a separate list on paper showing only the available lab tests. Helen is annoyed that she has to interrupt her work to change the lab test order. She opens the list of lab tests on the computer, verifies that she has now chosen the correct one by double-checking against the printed list, and then finishes the signing before going home 25 minutes later than planned.

Figure 2. An illustration of how one view of the health record for one patient can look in the system. There is one input line per date, and all inputs here are either notes or information about the patient contact made. The "Initials" column shows who made the input, and the "Sign." column shows the signing status: an empty space indicates that all signing has been completed; a single hashtag indicates that signing needs to be done; the overlapping hashtags (one or two) indicate that a post was signed but then changed and needs to be looked at and signed again.



Date	Type of input	Initials	Sign.
July 1, 2014	Note	HP	
July 1, 2014	Note	HP	
June 24, 2014	Note	JB	⧉#
June 23, 2014	Note	JB	⧉#
June 23, 2014	Contact	JB	⧉##
June 17, 2014	Note	CW	⧉##
June 17, 2014	Note	HP	#
June 17, 2014	Note	HP	#
June 17, 2014	Note	HP	#
June 17, 2014	Note	HP	#
June 17, 2014	Note	HP	#
June 17, 2014	Contact	HP	#

Figure 3. A different electronic health care record layout. This view presents information organized differently from the view shown in Figure 2; for example, the signing icons now appear in the center of the table and not to the right.



Date	Care contact type	Initials	Sign.	Unit	Note
July 8, 2014	Patient visit	HP	#	Administrative unit A	Written notes from patient contact.
July 7, 2014	Patient visit	JB		Administrative unit A	
July 4, 2014	Phone contact	JB	⧉##	Administrative unit A	
July 4, 2014	Note	CW	⧉##	Administrative unit A	Written notes from patient contact. Written notes from patient contact.
July 2, 2014	Patient visit	HP	#	Administrative unit A	
...					

Insights From the Scenarios

The scenarios have been synthesized from situations witnessed in health care practice to illustrate what working in a state-of-the-art EHR can be like. Several insights (Textbox 1) can be derived from these scenarios in order to understand how details in the design of and interaction with a digital system affect the work environment. The analysis of the scenarios leading to these insights is based on knowledge about how user-centered design processes and HCI design principles are aimed to guide the design of systems to avoid exactly the kind of situations demonstrated in the scenarios.

Textbox 1. Overview of insights derived from studying the usage of digital systems in contemporary health care practice.

- Insight 1: The digital system forces the user into inefficient workflows
- Insight 2: HCI design principles—independent of user, task, and context—are not adopted
- Insight 3: Workarounds become permanent solutions
- Insight 4: Different domains of ergonomics interact in creating the digital work environment

Insight 1: The Digital System Forces the User Into Inefficient Workflows

How different functions in the digital system are designed forces the user to adapt how tasks are performed. When the new EHR was introduced in the home health care organization (scenario 1), the communication routine changed. The design of the new system increased the number of messages that the user had to process each day. Similarly, the general practitioner at the primary care center (scenario 2) postponed the signing task since the system hindered rather than supported her in this process by not providing an intuitive workflow or a consistent design. This resulted in an inefficient procedure, causing both frustration and physical strain due to excessive clicking. Demands for more extensive signing in the primary care EHR (scenario 2) were deliberately introduced as a new work routine when implementing the new EHR in order to increase patient safety. Hence, the extra effort required to perform this task may be justified. However, how it was implemented in the system, with a confusing workflow, poor feedback, and inconsistency, could have been avoided by applying better knowledge of how the user works. In contrast, the increased message handling in home health care (scenario 1) was not deliberately introduced, and its effect on the time and effort needed to process the messages was not acknowledged or understood until the system came into practical use.

Insight 2: HCI Design Principles—Independent of User, Task, and Context—Are Not Adopted

A user-centered process is always desired, but even without such a process there are *fundamental HCI design principles* that should be adopted. A long list, as shown in Figure 1, is a long list regardless of the specific user and context. Offering a list that is not searchable or does not adapt to the user's most commonly chosen items is *inefficient* and can always be considered poor design. To increase usability, the list could have been either (1) sorted by intervention names instead of codes, (2) searchable by typing the first letters of the intervention name, or (3) sorted with the most commonly used interventions at the top of the list. The extensive clicking described in scenario 1 reveals a *complex information hierarchy* with many levels of information and interaction procedures that are not optimized for the task. *Inconsistency* in the system (in scenario 2), with different ways of presenting the signing information in one view compared with another (Figure 2, Figure 3), is also undesirable. Designing the user interface in this way is not in line with HCI design principles, where *efficiency*, *consistency*, and supportive *information structures* are examples of fundamental rules. Existing HCI design principles are based on information about human cognitive processes; by contravening these principles, interactions with the system occupy cognitive resources that could be used for better purposes [28-30,43].

Insight 3: Workarounds Become Permanent Solutions

In scenario 1, the nurse must fake a printing procedure to create a PDF that presents the patient's health record in the desired chronological order. This solution works around the current default, which presents the information in reverse chronological order, not in line with how the nurse reads the information. It is not technically difficult to implement this correction, so one wonders why it was implemented incorrectly from the beginning. This could have happened because of a lack of understanding of the user's tasks, revealing that a user-centered processes did not guide the system development. Since redesigning the software entails additional costs, a so-called workaround solution—creating a PDF—is offered. Likewise, the medical laboratory assistant in scenario 2 had created a workaround solution for handling the long list of lab tests that caused many incorrect orders. This solution involved a printed paper with a shorter version of the list, including only the available lab tests. These examples of workaround solutions are indicative of faulty system implementation and should lead to changes in the digital system. Instead, it is the users who must adapt, and the risk is that such workarounds can become a standard part of the interaction [3].

Insight 4: Different Domains of Ergonomics Interact in Creating the Digital Work Environment

Although HCI is traditionally associated mainly with the domain of cognitive ergonomics, we would like to emphasize that the development, implementation, and usage of a digital system entails interaction between several ergonomic domains: physical, cognitive, and organizational ergonomics [44]. This view is in line with what is called mesoergonomics, combining microergonomics, represented by physical and cognitive ergonomics at the individual level, and macroergonomics, represented by organizational aspects at the sociotechnical system level [45].

In the scenarios, we see examples of the users having to traverse long lists of information, interpret vague feedback, wait for the system to load, and be interrupted in tasks, all of which are related to cognitive ergonomics. Time-consuming or cumbersome interactions are frustrating and tiring to deal with and increase the cognitive load [43]. Delays of longer than one second in the home health care EHRs are a problem since time is a main determinant of cognitive load and mental effort [46].

Physical ergonomics is also part of the interaction. A design that involves small buttons, many clicks, and an inefficient layout, as in the two scenarios, can cause strain, mainly in the neck and shoulders [47,48]. The scenarios also present examples related to visual ergonomics, such as tiny user interface components, which may cause eye strain [49].

How tasks are performed in the EHR system is closely connected to, for example, how work is performed and how staff interact and communicate. We saw examples of this in scenario 1, in how communication was performed through messaging in home health care, and in scenario 2, when the physician at the primary care center had to allot extra time for administration in order to handle the new way of signing posts in the system. Hence, organizational aspects of ergonomics are also part of the digital work environment [5,45].

Analysis of the Four Insights

What the four above insights share is that none of them is really novel: they all concern issues that have long been known, and in a sense solved, by HCI and ergonomics researchers. The insights were extracted from observations of real use situations involving modern EHRs and typical everyday work by health care professionals, meaning that the related issues are still very real for the typical users of such systems. Furthermore, several recent scientific publications support the insights noted in the scenarios, and we conclude that usability issues are still present, workarounds are still common, and there are no indications that these issues are decreasing in prevalence [3,5,22,24,50]. The systems, furthermore, introduce *new* tasks that must be performed in relation to the digital system [51], potentially making the work situation more complex rather than more efficient.

Concerning the first insight about introducing *inefficient workflows*, it would have been valuable if the practical implications of the new functionalities had been better understood before the implementation. This could partly be due to *lack of user studies and user involvement* in the design process [27]. Another factor could be the *lack of procurement requirements* regarding nonfunctional aspects, such as workflow, being formulated during system procurement [52]. This indicates a need for support when formulating requirements in the procurement process, to ensure both adequate user involvement and consideration of current workflows and work practice demands. Hence, requirements engineering and procurement strategies that ensure good usability and ergonomically sound systems clearly need more attention.

The second insight about violating *HCI design principles* differs in that the problem and, in principle, the solution are very clear. HCI design principles should be followed, and there are methods such as heuristic evaluation and cognitive walkthrough that can help designers and developers determine whether they have been implemented correctly [29,53]. Again, part of the issue is that these principles are insufficiently highlighted in the procurement process. Another aspect of the issue is that it can be hard for users to determine whether HCI design principles are violated in existing systems, so they are easy to miss when procuring new systems. While there are tools for developers and designers to evaluate this, they might not always be suitable for users and their organizations.

The third insight about *workarounds that become permanent solutions* has origins similar to those of the first and second insights—namely, their causes should be sought in the procurement process and in the evaluation of existing systems. Workarounds exist because the practical implications of new functionalities were not considered when the system was designed. Later, when the system is implemented in practice, tools are lacking for evaluating the system and pinpointing how the workarounds negatively affect practice. Tools and methods for evaluating digital systems that can be used in practice are important, as they *legitimize* the identification of issues with digital systems that affect health care practice.

The fourth and last insight concerns the interaction between *different domains of ergonomics*. While HCI theory is largely based on cognitive ergonomics, both physical and organizational ergonomics are involved when introducing new digital tools. As noted above, the EHR design used in scenario 2 actually forced the physician to *reorganize* her work to manage the signing task. Regarding physical ergonomics, using established HCI design principles such as Fitts' law [54] to evaluate buttons and guide the layout and using visual ergonomics guidelines [55] on, for example, font size and color schemes, would be good starting points [56].

As the ongoing digitalization introduces new ways of working, including new hardware (eg, mobile documentation on tablets or wearable sensors) as part of the work equipment, interaction between different ergonomic domains can be expected to become more complex. A few relevant studies have already emerged. Johnson et al [57] identified a risk of musculoskeletal injuries when exchanging mouse and keyboard for a wearable sensor arm band. Xue [58] reviewed use areas and possible risks of wearables and also identified musculoskeletal strain as a possible risk.

Paths Forward

Overview

Based on the above discussion, we identify three paths toward a more sustainable digital work environment. The first path focuses on providing HCI theory–based *tools usable in health care practice to evaluate digital systems* as they are used in the field. The second path focuses on specifying *generic requirement formulations* related to usability, for use in the procurement process. The third path advocates introducing *digital ergonomics* as an embracing concept to emphasize all the ergonomic aspects involved in the digitalization process.

Path 1: Tools for Evaluating Digital Systems in the Field

Studies exploring system usability are typically based on the users' subjective grading of various usability statements related to the system (eg, [22,50]). This lets us know that users are sometimes frustrated when using digital systems. However, the specific reasons why a system is considered good or bad are not clearly identified: Is it because the design violates HCI design principles, because the workflow does not match the user's task, because the system's content layout does not fit the size and shape of the screen, or simply because the system is technically unstable? Furthermore, it is unreasonable to expect users in health care practice to identify HCI and ergonomic issues. In contrast, ergonomists, human factor specialists, and others responsible for the work environment should be able to do this with the right tools. They already perform risk assessment using ergonomic tools that, for example, measure lighting, musculoskeletal strain, and cognitive load, tools developed from a vast knowledge base in the related scientific subfield and adapted to be practically applicable. Some attempts are being made to support field assessment of the digital work environment [59,60], but we lack mature tools applicable *in practice by practitioners*. Revising existing principles and guidelines from HCI theory is required to transform these from theoretical, highly specialized methods into practically applicable tools for persons who are not HCI experts [61]. With such tools, it will be possible to evaluate an existing system and categorize and prioritize among the identified issues. The results of such an evaluation can then be used either to learn about the current digital work environment and make improvements based on this, or as a source of information when procuring new systems.

Path 2: Generic Requirement Formulations

When procuring a new digital system, such as an EHR, requirements related to HCI and ergonomics are often either completely lacking or expressed in unverifiable terms [7]. Borg et al [62] showed that, in procuring digital tools, authorities primarily focus on functional requirements (what the system should do) rather than on quality demands (how well the functions are supported) such as usability. On the software development side, functional requirements are prioritized over quality-related requirements [63]. It has also been shown that requirements are lost during the software development process due to communication gaps [64]. There is apparently a need for support when it comes to formulating distinct and verifiable requirements for usability and ergonomics.

A lack of adequate *systems acquisition competence* is often identified as an explanation for bad digital systems being acquired [65]. It has even been suggested that health care professionals themselves should be more active in developing digital systems by, for example, taking responsibility for managing and designing the systems [66]. However, health care professionals are not, and should not need to be, specialists in software development, ergonomics, or design. Furthermore, it is unreasonable to expect all organizations acquiring digital systems—ranging from small home health care units and local primary care centers to large regional hospitals—to possess the extensive competence that is required to do this.

By creating guidelines for procurement that include user involvement, together with *generic requirement formulations* concerning usability and ergonomics, which can basically be taken “off the shelf,” it becomes easier to address those types of needs when specifying requirements. In this way, the risk of overlooking these aspects or losing track of them along the way decreases. Currently, no such generic requirement formulations exist; they need to be created from the current knowledge base regarding different aspects of ergonomics.

Path 3: Digital Ergonomics—A Comprehensive Ergonomics Approach to Digitalization

As mentioned above, ergonomics has many aspects, and frameworks for combining these aspects have been proposed to allow for the simultaneous achievement of individual employee and organizational goals [45]. Digitalization has transformed work life and is itself affected by all levels of ergonomics. The examples from the studied scenarios indicate that ergonomic aspects are not fully considered in the digitalization process, resulting in poorly designed systems with suboptimized workflows causing unnecessary physical and mental strain. Increasing the awareness of ergonomic aspects in the digitalization process is necessary, and applying a holistic systems ergonomics perspective is desired to capture the complexity surrounding the digital work environment.

To meet this need, we propose the introduction of *digital ergonomics* as a concept used to emphasize the importance of adopting a comprehensive ergonomics approach, including relevant aspects from each subdomain of ergonomics, to develop sustainable digital work environments. We propose the following definition: *Digital ergonomics is the multidisciplinary science concerned with the application of theory, principles, data, and methods to the design of digital systems and the digital work environment, in order to optimize human well-being and overall system performance.*

This definition is based on the International Ergonomics Association’s [67] definition of ergonomics, adapted to narrow the focus on the implementation and usage of digital systems at work. Digital systems and their surroundings are arguably already part of ergonomics, as ergonomics as a discipline is rooted in sociotechnical values. However, since both HCI and ergonomic issues continue to emerge in health care practices and other work practices, an effort to direct attention to this area is needed. Uniting around the *digital ergonomics* concept would help users and developers focus on and clarify the required conditions for developing knowledge, tools, and methods for systematically addressing the work environment in relation to digitalization.

Conclusion

Discussion of the digitalization of work is often general and insufficiently detailed to capture the full complexity of digitalization, which includes everything from effects on efficiency, safety, ethics, power relations, the work environment, and new ways of organizing work to designing the technology itself. Much research into digitalization emphasizes how health care is being transformed as a practice. As demonstrated in the above scenarios, many problems associated with digitalization only appear in actual use situations, where they also become issues for individual workers. Furthermore, many of these problems are completely unnecessary, since they are not novel and knowledge of how to avoid them has long existed.

In order to bridge this gap, initiatives are needed to more actively integrate ergonomics and HCI theory into practice. We suggest three paths forward so as not to overlook fundamental aspects of implementing and using digital systems in order to create a sustainable digital work environment: (1) development of *tools for evaluating the digital work environment in the field* capable of capturing details of the actual usage of digital systems, based on HCI theory, while being applicable in practice; (2) *generic requirement formulations* to ensure that demands related to all aspects of the digital work environment are considered in the procurement process and not lost sight of along the way; and (3) the introduction of the concept *digital ergonomics* in order to extract and merge aspects of all ergonomics domains that may influence the digital work environment. These three concrete suggestions would help bridge the gap between research-based knowledge and effects in practice in order to achieve sustainable digital work environments.

Conflicts of Interest

None declared.

References

1. Koster J, Stewart E, Kolker E. Health Care Transformation: A Strategy Rooted in Data and Analytics. Acad Med 2016 Feb;91(2):165-167. [doi: 10.1097/ACM.0000000000001047] [Medline: 26630610]
2. Agarwal R, Gao G, DesRoches C, Jha AK. The Digital Transformation of Healthcare: Current Status and the Road Ahead. Information Systems Research 2010 Dec;21(4):796-809. [doi: 10.1287/isre.1100.0327]
3. Lawler EK, Hedge A, Pavlovic-Veselinovic S. Cognitive ergonomics, socio-technical systems, and the impact of healthcare information technologies. International Journal of Industrial Ergonomics 2011 Jul;41(4):336-344. [doi: 10.1016/j.ergon.2011.02.006]
4. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc 2011 May 01;18(3):327-334 [FREE Full text] [doi: 10.1136/amiajnl-2011-000094] [Medline: 21422100]
5. Carayon P, Hoonakker P. Human Factors and Usability for Health Information Technology: Old and New Challenges. Yearb Med Inform 2019 Aug;28(1):71-77 [FREE Full text] [doi: 10.1055/s-0039-1677907] [Medline: 31419818]
6. Verbeek PP. Morality in Design: Design Ethics and the Morality of Technological Artifacts. In: Philosophy and Design: From Engineering to Architecture. Dordrecht: Springer Netherlands; 2008:91-103.
7. Sandblad B, Gulliksen J, Åborg C, Boivie I, Persson J, Göransson B, et al. Work environment and computer systems development. Behaviour & Information Technology 2003 Nov;22(6):375-387. [doi: 10.1080/01449290310001624356]
8. Griffiths KL, Mackey MG, Adamson BJ. The impact of a computerized work environment on professional occupational groups and behavioural and physiological risk factors for musculoskeletal symptoms: a literature review. J Occup Rehabil 2007 Dec;17(4):743-765. [doi: 10.1007/s10926-007-9108-x] [Medline: 17987369]
9. Shortliffe EH. Computer programs to support clinical decision making. JAMA 1987 Jul 03;258(1):61-66. [Medline: 3586293]
10. Andreoli K, Musser LA. Computers in Nursing Care: The State of the Art. In: Saba VK, Rieder KA, Pocklington DB, editors. Nursing and Computers: An Anthology. New York, NY: Springer New York; 1989:176-186.
11. Arnetz BB. Technological stress: psychophysiological aspects of working with modern information technology. Scand J Work Environ Health 1997;23 Suppl 3:97-103 [FREE Full text] [Medline: 9456075]
12. Malone TW. How do people organize their desks? Implications for the design of office information systems. ACM Trans. Inf. Syst 1983 Jan;1(1):99-112. [doi: 10.1145/357423.357430]
13. Neale DC, Carroll JM, Rosson MB. Evaluating computer-supported cooperative work: models and frameworks. 2004 Presented at: 2004 ACM conference on Computer supported cooperative work; 2004; Chicago, Illinois, USA p. 112-121. [doi: 10.1145/1031607.1031626]
14. Grudin J. Computer-supported cooperative work: history and focus. Computer 1994 May;27(5):19-26. [doi: 10.1109/2.291294]
15. Baig MM, GholamHosseini H, Connolly MJ. Mobile healthcare applications: system design review, critical issues and challenges. Australas Phys Eng Sci Med 2015 Mar;38(1):23-38. [doi: 10.1007/s13246-014-0315-4] [Medline: 25476753]
16. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
17. International Organization for Standardization (ISO). Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. Geneva: ISO; 2018.
18. Cho I, Choi WJ, Choi WH. Usability Testing and Comparison of Six Electronic Nursing Record Systems. CIN: Computers, Informatics, Nursing 2014;32(8):364. [doi: 10.1097/01.ncn.0000453182.72347.7c]
19. Dowding D, Merrill JA, Barrón Y, Onorato N, Jonas K, Russell D. Usability Evaluation of a Dashboard for Home Care Nurses. Comput Inform Nurs 2019 Jan;37(1):11-19 [FREE Full text] [doi: 10.1097/CIN.0000000000000484] [Medline: 30394879]
20. Witteman HO, Dansokho SC, Colquhoun H, Coulter A, Dugas M, Fagerlin A, et al. User-centered design and the development of patient decision aids: protocol for a systematic review. Syst Rev 2015;4:11 [FREE Full text] [doi: 10.1186/2046-4053-4-11] [Medline: 25623074]

21. Liljegren E. Usability in a medical technology context assessment of methods for usability evaluation of medical equipment. International Journal of Industrial Ergonomics 2006 Apr;36(4):345-352. [doi: 10.1016/j.ergon.2005.10.004]
22. Bundschuh BB, Majeed RW, Bürkle T, Kuhn K, Sax U, Seggewies C, et al. Quality of human-computer interaction-results of a national usability survey of hospital-IT in Germany. BMC Med Inform Decis Mak 2011 Nov 09;11:69 [FREE Full text] [doi: 10.1186/1472-6947-11-69] [Medline: 22070880]
23. Beuscart-Zéphir MC, Brender J, Beuscart R, Ménager-Depriester I. Cognitive evaluation: how to assess the usability of information technology in healthcare. Comput Methods Programs Biomed 1997 Sep;54(1-2):19-28. [doi: 10.1016/s0169-2607(97)00030-8] [Medline: 9290916]
24. Viitanen J, Hyppönen H, Lääveri T, Vänskä J, Reponen J, Winblad I. National questionnaire study on clinical ICT systems proofs: physicians suffer from poor usability. Int J Med Inform 2011 Oct;80(10):708-725. [doi: 10.1016/j.ijmedinf.2011.06.010] [Medline: 21784701]
25. Lowery JC, Martin JB. Evaluation of healthcare software from a usability perspective. J Med Syst 1990 Apr;14(1-2):17-29. [doi: 10.1007/BF00995877] [Medline: 2373969]
26. Karsh BT. Beyond usability: designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394 [FREE Full text] [doi: 10.1136/qhc.13.5.388] [Medline: 15465944]
27. Gould JD, Lewis C. Designing for usability: key principles and what designers think. Commun. ACM 1985;28(3):300-311. [doi: 10.1145/3166.3170]
28. Norman D. The design of everyday things. New York, NY: Basic Books; 2013.
29. Nielsen J, Molich R. Heuristic evaluation of user interfaces. 1990 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 1990; Seattle, Washington, USA p. 249-256. [doi: 10.1145/97243.97281]
30. Schneiderman B. Designing the user interface: Strategies for effective human-computer interaction. Reading, MA: Addison-Wesley; 1987.
31. International Organization for Standardization (ISO). Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. Geneva: ISO; 2019.
32. Gulliksen J, Boivie I, Göransson B. Usability professionals—current practices and future development. Interacting with Computers 2006 Jul;18(4):568-600. [doi: 10.1016/j.intcom.2005.10.005]
33. Ehn P. Neither Bauhäusler nor nerd - educating the interaction designer. 2002 Presented at: 4th conference on Designing interactive systems: processes, practices, methods, and techniques; 2002; London, England p. 19-23. [doi: 10.1145/778712.778717]
34. Meloncon L, Henschel S. Current state of U.S. undergraduate degree programs in technical and professional communication. Technical Communication 2013;60(1):45-64.
35. Kaipio J, Lääveri T, Hyppönen H, Vainiomäki S, Reponen J, Kushniruk A, et al. Usability problems do not heal by themselves: National survey on physicians' experiences with EHRs in Finland. Int J Med Inform 2017 Jan;97:266-281 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.10.010] [Medline: 27919385]
36. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
37. Rydenfält C, Persson J, Erlingsdottir G, Johansson G. eHealth Services in the Near and Distant Future in Swedish Home Care Nursing. Comput Inform Nurs 2019 Jul;37(7):366-372. [doi: 10.1097/CIN.0000000000000536] [Medline: 31135467]
38. Lareau D. The Other Health Care Crisis: The Lack Of EHR Usability.: Forbes; 2020. URL: https://www.forbes.com/sites/forbestechcouncil/2020/09/21/the-other-health-care-crisis-the-lack-of-ehr-usability/?sh=66beef0c293e [accessed 2021-02-15]
39. Miliard M. Mayo Clinic study links EHR usability with clinician burnout.: Healthcare IT News; 2019. URL: https://www.healthcareitnews.com/news/mayo-clinic-study-links-ehr-usability-clinician-burnout [accessed 2021-02-15]
40. Cajander Å, Lárusdóttir MK, Lind T, Nauwerck G. Walking in the jungle with a machete: ICT leaders’ perspectives on user-centred systems design. Behaviour & Information Technology 2021 Jan 18:1-15. [doi: 10.1080/0144929x.2020.1864776]
41. Rydenfält C, Persson J. The usability and digitalization of healthcare. XRDS 2020 Apr 02;26(3):42-45. [doi: 10.1145/3383386]
42. Carayon P, Salwei ME. Moving toward a sociotechnical systems approach to continuous health information technology design: the path forward for improving electronic health record usability and reducing clinician burnout. J Am Med Inform Assoc 2021 Apr 23;28(5):1026-1028. [doi: 10.1093/jamia/ocab002] [Medline: 33537756]
43. Wisner K, Lyndon A, Chesla CA. The electronic health record's impact on nurses' cognitive work: An integrative review. Int J Nurs Stud 2019 Jun;94:74-84. [doi: 10.1016/j.ijnurstu.2019.03.003] [Medline: 30939418]
44. Wilson JR, Sharples S. Evaluation of Human Work. Boca Raton, FL: CRC Press; 2015.
45. Karsh BT, Waterson P, Holden RJ. Crossing levels in systems ergonomics: a framework to support 'mesoergonomic' inquiry. Appl Ergon 2014 Jan;45(1):45-54 [FREE Full text] [doi: 10.1016/j.apergo.2013.04.021] [Medline: 23706573]
46. Barrouillet P, Bernardin S, Portrat S, Vergauwe E, Camos V. Time and cognitive load in working memory. J Exp Psychol Learn Mem Cogn 2007 May;33(3):570-585. [doi: 10.1037/0278-7393.33.3.570] [Medline: 17470006]
47. Arvidsson I. Musculoskeletal disorders in demanding computer work - with air traffic control as a model. Lund, Sweden: Lund University; 2008.

48. Laursen B, Jensen BR, Garde AH, Jørgensen AH. Effect of mental and physical demands on muscular activity during the use of a computer mouse and a keyboard. Scand J Work Environ Health 2002 Aug;28(4):215-221 [FREE Full text] [doi: 10.5271/sjweh.668] [Medline: 12199422]
49. Rosenfield M. Computer vision syndrome (a.k.a. digital eye strain). Optometry in Practice 2016;17(1):1-10.
50. Kaipio J, Hyppönen H, Lääveri T. Physicians' Experiences on EHR Usability: A Time Series from 2010, 2014 and 2017. Stud Health Technol Inform 2019;257:194-199. [Medline: 30741195]
51. Golay D. An invisible burden: An experience-based approach to nurses' daily work life with healthcare information technology. Uppsala, Sweden: Uppsala University; 2019.
52. Ullah S, Iqbal M, Khan AM. A survey on issues in non-functional requirements elicitation. 2011 Presented at: International Conference on Computer Networks and Information Technology; 2011; Abottabad, Pakistan. [doi: 10.1109/iccnit.2011.6020890]
53. Wharton C, Rieman J, Lewis C, Polson P. The cognitive walkthrough method: a practitioner's guide. In: Nielsen J, Mack RL, editors. Usability inspection methods. New York, NY: John Wiley & Sons, Inc; 1994:105-140.
54. Fitts PM. The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 1954 Jun;47(6):381-391. [Medline: 13174710]
55. Long J. What is visual ergonomics? Work 2014 Jan 01;47(3):287-289. [doi: 10.3233/WOR-141823] [Medline: 24448008]
56. Rello L, Pielot M, Marcos MC. Make it big! The effect of font size and line spacing on online readability. 2016 Presented at: CHI conference on Human Factors in Computing Systems; 2016; San Jose California, USA. [doi: 10.1145/2858036.2858204]
57. Johnson M, Conrardy B, Kohama Z, Piper A. Repetitive upper extremity musculoskeletal risks utilizing wearable sensor arm band versus keyboard and mouse for input. 2017 Presented at: Institute of Industrial & Systems Engineers (IISE) Annual Conference; 2017; Pittsburgh, PA.
58. Xue Y. A review on intelligent wearables: Uses and risks. Human Behav and Emerg Tech 2019 Sep 16;1(4):287-294 [FREE Full text] [doi: 10.1002/hbe2.173]
59. Kushniruk A, Senathirajah Y, Borycki E. Effective usability engineering in healthcare: A vision of usable and safer healthcare IT. : International Medical Informatics Association (IMIA) and IOS Press; 2017 Presented at: MEDINFO 2017: Precision Healthcare through Informatics; 2017; Hangzhou, China p. 1066-1069.
60. Smith SW, Koppel R. Healthcare information technology's relativity problems: a typology of how patients' physical reality, clinicians' mental models, and healthcare information technology differ. J Am Med Inform Assoc 2014;21(1):117-131 [FREE Full text] [doi: 10.1136/amiajnl-2012-001419] [Medline: 23800960]
61. Colusso L, Bennett CL, Hsieh G, Munson SA. Translational resources: Reducing the gap between academic research and HCI practice. : ACM; 2017 Presented at: Conference on Designing Interactive Systems; 2017; Edinburgh, United Kingdom p. 957-968. [doi: 10.1145/3064663.3064667]
62. Borg M, Olsson T, Franke U, Assar S. Digitalization of Swedish government agencies: A perspective through the lens of a software development census. 2018 Presented at: International Conference on Software Engineering: Software Engineering in Society; 2018; Gothenburg, Sweden p. 37-46. [doi: 10.1145/3183428.3183434]
63. Ebert A, Humayoun SR, Seyff N, Perini A, Barbosa SDJ. Bridging the gap between requirements engineering and human-computer interaction. : Springer International Publishing; 2016 Presented at: UsARE: International Workshop on Usability- and Accessibility-Focused Requirements Engineering; 2014; Karlskrona, Sweden p. 3-7. [doi: 10.1007/978-3-319-45916-5_1]
64. Bjarnason E, Wnuk K, Regnell B. 2011 Presented at: IEEE 19th International Requirements Engineering Conference; 2011; Trento, Italy. [doi: 10.1109/re.2011.6051639]
65. Thomsen M. Beställarkompetens vid upphandling och utveckling av IT? Om kompetensframväxt i skuggan av kunskapsfragmentering (Aquisition competence in procurement and development of IT). Lund, Sweden: Lund University; 2010.
66. While A, Dewsbury G. Nursing and information and communication technology (ICT): a discussion of trends and future directions. Int J Nurs Stud 2011 Oct;48(10):1302-1310. [doi: 10.1016/j.ijnurstu.2011.02.020] [Medline: 21474135]
67. What is ergonomics?.: International Ergonomics Association (IEA); 2020. URL: https://iea.cc/what-is-ergonomics/ [accessed 2021-08-20]

Abbreviations

EHR: electronic health care record
HCI: human-computer interaction

Edited by R Kukafka; submitted 22.12.20; peer-reviewed by A Jatoba, B Smaradottir; comments to author 07.02.21; revised version received 19.02.21; accepted 24.05.21; published 22.06.21.

Please cite as:
Persson J, Rydenfält C
Why Are Digital Health Care Systems Still Poorly Designed, and Why Is Health Care Practice Not Asking for More? Three Paths Toward a Sustainable Digital Work Environment
J Med Internet Res 2021;23(6):e26694
URL: https://www.jmir.org/2021/6/e26694
doi:10.2196/26694
PMID:34156336

Original Paper

Sexual Health Influencer Distribution of HIV/Syphilis Self-Tests Among Men Who Have Sex With Men in China: Secondary Analysis to Inform Community-Based Interventions

Nancy Yang[1,2,3,4*], MD; Dan Wu[3,5*], PhD; Yi Zhou[6*], MPH; Shanzi Huang[6], MPH; Xi He[7], BA; Joseph Tucker[3,5], MD, PhD; Xiaofeng Li[6], MPH; Kumi M Smith[8], PhD; Xiaohui Jiang[6]; Yehua Wang[3], MPH; Wenting Huang[3], MSc; Hongyun Fu[9], PhD; Huanyu Bao[3], MSc; Hongbo Jiang[10], PhD; Wencan Dai[6], PhD; Weiming Tang[2,3,4], PhD

[1]University of Hawai'i John A. Burns School of Medicine, Hawai'i, HI, United States
[2]Dermatology Hospital of Southern Medical University, Guangzhou, China
[3]University of North Carolina Project-China, Guangzhou, China
[4]Institute of Global Health and STI Research of Southern Medical University, Guangzhou, China
[5]Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, United Kingdom
[6]Zhuhai Center for Diseases Control and Prevention, Zhuhai, China
[7]Zhuhai Xutong Voluntary Services Center, Zhuhai, China
[8]Division of Epidemiology and Community Health, The University of Minnesota Twin cities, Minneapolis, MN, United States
[9]Division of Community Health and Research, Eastern Virginia Medical School, Norfolk, VA, United States
[10]Department of Epidemiology and Biostatistics, School of Public Health, Guangdong Pharmaceutical University, Guangzhou, China
[*]these authors contributed equally

Corresponding Author:
Weiming Tang, PhD
Dermatology Hospital of Southern Medical University
No 2 Lujing Road
Yuexiu District
Guangzhou, 510091
China
Phone: 86 15920567132
Email: weiming_tang@med.unc.edu

Abstract

Background: Social network–based strategies can expand HIV/syphilis self-tests among men who have sex with men (MSM). Sexual health influencers are individuals who are particularly capable of spreading information about HIV and other sexually transmitted infections (STIs) within their social networks. However, it remains unknown whether a sexual health influencer can encourage their peers to self-test for HIV/syphilis.

Objective: The aims of this study were to examine the impact of MSM sexual health influencers on improving HIV/syphilis self-test uptake within their social networks compared to that of nonsexual health influencers.

Methods: In Zhuhai, China, men 16 years or older, born biologically male, who reported ever having had sex with a man, and applying for HIV/syphilis self-tests were enrolled online as indexes and encouraged to distribute self-tests to individuals (alters) in their social network. Indexes scoring >3 on a sexual health influencer scale were considered to be sexual health influencers (Cronbach α=.87). The primary outcome was the mean number of alters encouraged to test per index for sexual health influencers compared with the number encouraged by noninfluencers.

Results: Participants included 371 indexes and 278 alters. Among indexes, 77 (20.8%) were sexual health influencers and 294 (79.2%) were noninfluencers. On average, each sexual health influencer successfully encouraged 1.66 alters to self-test compared to 0.51 alters encouraged by each noninfluencer (adjusted rate ratio 2.07, 95% CI 1.59-2.69). More sexual health influencers disclosed their sexual orientation (80.5% vs 67.3%, *P*=.02) and were community-based organization volunteers (18.2% vs 2.7%, *P*<.001) than noninfluencers. More alters of sexual health influencers came from a rural area (45.5% vs 23.8%, *P*<.001), had below-college education (57.7% vs 37.1%, *P*<.001), and had multiple casual male sexual partners in the past 6 months (25.2% vs 11.9%, *P*<.001).

Conclusions: Being a sexual health influencer was associated with encouraging more alters with less testing access to self-test for HIV/syphilis. Sexual health influencers can be engaged as seeds to expand HIV/syphilis testing coverage.

(J Med Internet Res 2021;23(6):e24303) doi:10.2196/24303

KEYWORDS

sexual health influencer; men who have sex with men; HIV; syphilis; self-test; sexual health; influencer; social network; peers

Introduction

A social network is a network of individuals connected by interpersonal relationships [1]. Social network–based interventions are promising for promoting HIV testing and sexual health [2,3], including among men who have sex with men (MSM) who are at higher risk of HIV and other sexually transmitted infections (STIs) [4-10]. Social network–based strategies can increase testing access [11-14] and mitigate the stigma preventing MSM from seeking health care [10,15]. Many studies have found that individuals at the center of their networks can encourage health behavior change among their peers [16]. However, there is limited research on whether social influencers are better than noninfluencers at promoting HIV testing. A previous study found that MSM with more sexual health influence had greater engagement in sexual health campaigns and adoption of HIV/syphilis testing [17], but did not examine the interaction between influential MSM and their social network.

Strategies are needed to identify influential individuals for effective network dissemination of HIV/STI interventions. We define "sexual health influencers" as individuals whose sexual health knowledge and behaviors are more likely to influence, than be influenced by, peers in their social network, based on a prior MSM sexual influence study in China [17]. Although the precise criteria for "peer" differs among individual studies, peers are generally understood to be individuals that share key characteristics such as sexual orientation [18]. In this context, we consider sexual health influencers to be a subset of peers that are selected as peer educators based on their preexisting influence and social ties. Sexual health influencers are different from popular opinion leaders (POLs), a commonly used model in peer education interventions. POLs are rigorously trained for influencing their target audience, with whom they may not have prior social ties even if they are peers [2]. There is evidence that the effectiveness of POLs [4-8] is predicated on preexisting influence, and POL interventions that do not consider preexisting influence are less likely to succeed [19]. Additionally, interventions that do not engage influential individuals for dissemination may exclude hard-to-reach individuals [3] such as those with limited health care knowledge and access, and those who do not identify with at-risk communities but practice at-risk behaviors.

Our previous implementation program used a secondary distribution strategy to promote HIV/syphilis self-testing among Chinese MSM [20]. Secondary distribution is a social network–based strategy that involves giving one individual (index) multiple self-testing kits for distribution to their social contacts (alters) [12]. Evaluation of this implementation program demonstrated that secondary distribution can expand HIV/syphilis self-testing among MSM in a middle-income country [20]. Nonetheless, it remains unexplored whether sexual health influencer status is associated with greater promotion of HIV/STI self-testing among peers, especially peers who have less access to testing.

The primary objective of this study was to evaluate whether sexual health influencer MSM are more effective at encouraging HIV/STI self-testing among their peers compared to noninfluencers. Additionally, we aimed to test the hypothesis that sexual health influencers can reach alters in greater need of HIV/STI testing than noninfluencers. Our findings can improve network-targeting strategies for distributing HIV and syphilis self-tests, and increase access to sexual health interventions for MSM.

Methods

Participants and Enrollment

This was a retrospective cohort analysis of data obtained from an implementation study performed in Zhuhai, China. Detailed methods were described previously [20]. We partnered with Zhuhai Xutong MSM Service Center (hereafter "Xutong"), a gay community-based organization (CBO) based in Zhuhai, China, to use their HIV/syphilis self-test (hereafter "self-test") distribution platform on WeChat (China's largest social media platform). Men who applied for self-tests on this platform were invited to participate in our study. Participants were enrolled as indexes if they were 16 years or older, born biologically male, ever had sex with a man, applied for at least one self-test during the study period, and willing to complete a follow-up survey. Following provision of informed consent online, indexes completed a baseline survey, and then provided up to five self-tests per application for a deposit of US $14.70 per self-test. Multiple applications were allowed. Indexes were encouraged to use the self-tests not only for themselves but to also distribute self-tests to individuals in their social networks (hereafter "alters"). Each self-test contained a unique QR code for anonymous upload of a test result photograph and a follow-up survey. Upon upload of results, the tester was given a US $3.00 incentive, and the deposit associated with that self-test was refunded to the index through WeChat. Additionally, alters were asked to report their self-test experience. Alters of all sexes, genders, and sexual orientation were included.

Prior to study implementation, the surveys were pilot-tested with a small group of representatives from our partner MSM community. Surveys and participant responses were stored on the secure survey platform Wenjuanxing (Sojump, Shanghai, China) protected with passcodes accessible only to the research team.

Identifying Sexual Health Influencers

Indexes were categorized as sexual health influencers or noninfluencers based on their responses to a 6-item sexual health influencer scale in the baseline survey. Each item was scored on a 5-point Likert-type scale, with a higher score indicating that the index is more likely to influence their peers and a lower score indicating that the index is more likely to be influenced by peers (Multimedia Appendix 1). Items were adapted from a scale previously studied in Chinese-speaking MSM populations [6,17]. Indexes were categorized as sexual health influencers if their mean score was greater than 3, using the same cut-off that previously identified sexual health influencers in a nationwide sample of MSM in China [17]. Cronbach α was .87 for the sexual health influencer scale in this study.

Survey Measures

We asked participants about their sociodemographic characteristics, including age, residence registration (rural or urban), sexual orientation, educational attainment, and monthly income. We also asked participants about their health behaviors, including disclosure of sexual orientation, number of male partners in the past 6 months, and prior HIV testing. For indexes, we identified those who were MSM CBO volunteers based on a list of volunteers from Xutong. For alters, we also asked whether they tested simultaneously with the index.

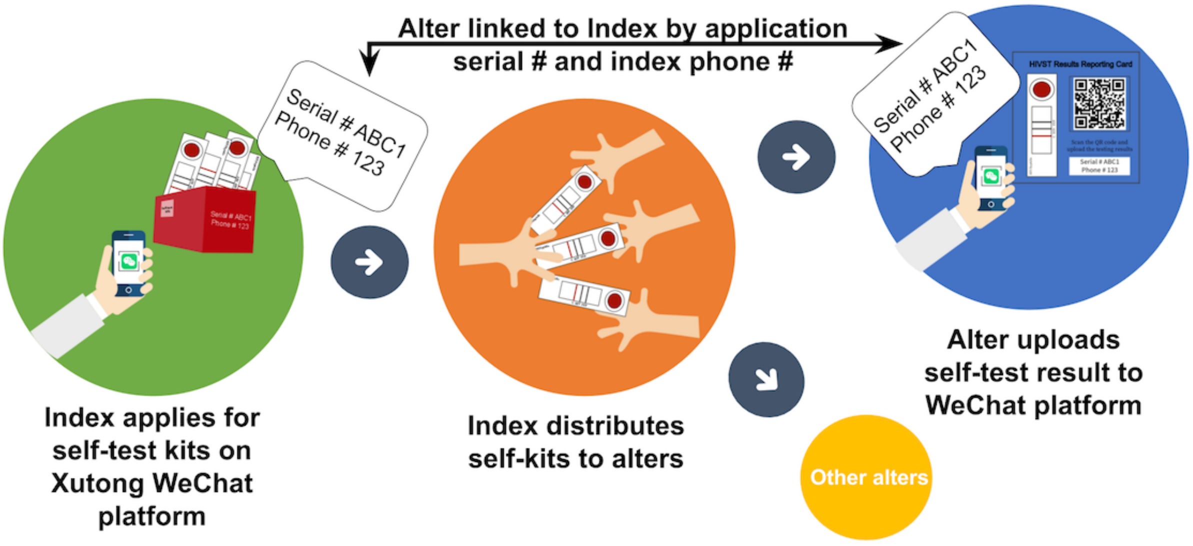
HIV and Syphilis Self-Test Results

All participants were asked to report their self-test results and upload a result photograph for verification. All results were reviewed by trained Xutong volunteers who followed up as needed for results verification or linkage to care. Only newly positive cases were counted in this study.

Linkage of Indexes and Alters

Each applying index had to report their phone number and was assigned a unique application code. Participants uploading a self-test result were asked to report the test kit application phone number and code, which were used to link alters and indexes (Figure 1). Alters not linked to an enrolled index were excluded from analysis.

Figure 1. HIV/syphilis self-tests application, distribution, and linkage of alters to indexes.



Outcomes

The primary outcome was whether sexual health influencers could encourage more alters to self-test compared to noninfluencers. Between sexual health influencers and noninfluencers, we compared the mean number of applications submitted, self-tests obtained, and alters who reported a self-test result. We controlled for the following potential confounders: income, disclosed sexual orientation, CBO volunteer status, and prior HIV testing.

Secondary outcomes included the following: among sexual health influencers versus noninfluencers who distributed to at least one alter, we compared the mean number of alters who were first-time testers, performed simultaneous tests with the index, and alters with HIV-reactive or syphilis-reactive tests. Volunteer status of the index was selected a priori as a confounder. We also compared the characteristics of alters who received a self-test from sexual health influencers versus noninfluencers.

Statistical Analysis

Descriptive analyses were performed for the sociodemographic and behavioral characteristics of the index and alter participants. The characteristics of sexual health influencers and noninfluencers, as well as characteristics of their respective alters, were compared using *t* tests and χ^2 tests. Poisson regression was used to estimate the ratio of distribution by sexual health influencers versus noninfluencers, which are reported as the adjusted rate ratio (aRR) and 95% CI. Additional variables were added to the regression to control for confounders. Statistical analyses were performed using SAS Version 9.4.

Ethical Statement

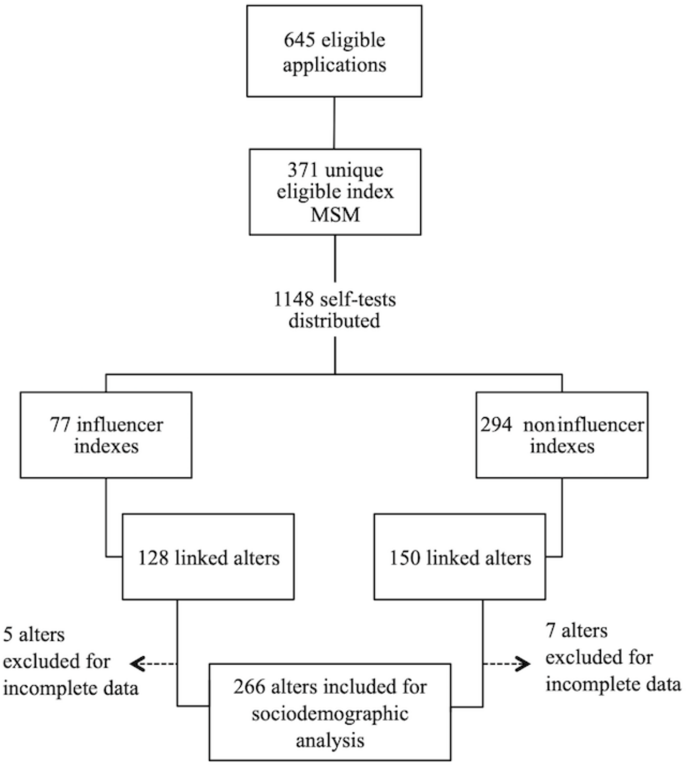
Prior to launching the study, ethical approval was obtained from the institutional review board at Zhuhai Municipal Center for Diseases Prevention and Control in China (ZHCDC2018014).

Results

Data Collection

Data were collected between June 17, 2018 and November 12, 2019. During this period, 371 unique indexes applied for 1148 self-tests, for which 1099 self-test results were returned by indexes and 278 unique alters linked to enrolled indexes. Of the alters, 266 completed the sociodemographic portion of the survey (Figure 2).

Figure 2. Flowchart of study participants.



Characteristics of Index Participants

Of the 371 indexes, 77 (20.8%) were sexual health influencers and 294 (79.2%) were noninfluencers. The distribution of sexual health influencer scores is provided in Multimedia Appendix 2. Characteristics of sexual health influencers and noninfluencers were largely similar but with some key differences. More sexual health influencers had disclosed their sexual orientation (80.5% vs 67.3%, *P*=.02) and were MSM CBO volunteers (18.2% vs 2.7%, *P*<.001) compared to noninfluencers. Additionally, more sexual health influencers had prior HIV testing compared to noninfluencers, although the difference was not statistically significant. Index participant characteristics are summarized in Table 1.

Table 1. Sociodemographic and health behavioral characteristics of the index participants in China, 2018-2019.

Characteristic	All indexes (N=371)	Sexual health influencers (n=77)	Noninfluencers (n=294)	*P* value
Age (years), mean (SD)	29 (6.9)	30 (8.2)	28 (6.5)	.06
Residence, n (%)				.79
Rural	164 (44.2)	33 (42.9)	131 (44.6)	
Urban	207 (55.8)	44 (57.1)	163 (55.4)	
Education, n (%)				.82
Less than college	173 (46.6)	35 (45.5)	138 (46.9)	
College or more	198 (53.4)	42 (54.6)	156 (53.1)	
Annual income (US $)[a], n (%)				.41
<8393	189 (50.9)	36 (46.8)	153 (52.0)	
≥8393	182 (49.1)	41 (53.3)	141 (48.0)	
Sexual orientation, n (%)				.62
Gay	259 (69.8)	52 (67.5)	207 (70.4)	
Bisexual/other	112 (30.2)	25 (32.5)	87 (29.6)	
Disclosed sexual orientation, n (%)				.02
Disclosed	260 (70.1)	62 (80.5)	198 (67.3)	
Undisclosed	111 (29.9)	15 (19.5)	96 (32.7)	
CBO[b] volunteer, n (%)				<.001
Yes	22 (5.9)	14 (18.2)	8 (2.7)	
No	394 (94.1)	63 (81.8)	286 (97.3)	
Ever tested for HIV, n (%)				.21
Yes	294 (79.3)	65 (84.4)	229 (77.9)	
No	77 (20.8)	12 (15.6)	65 (22.1)	
>1 casual male partner in past 6 months, n (%)				.31
Yes	117 (31.5)	28 (36.4)	89 (30.3)	
No	254 (68.5)	49 (63.6)	205 (69.7)	

[a] In 2017, the average annual income in China was US $7712 (National Bureau of Statistics China).

[b] CBO: community-based organization.

Rates of Index Self-Test Distribution and Results Return

On average, each sexual health influencer index had twice as many alters who returned a test result compared to each noninfluencer (aRR 2.07, 95% CI 1.59-2.69). Each sexual health influencer also submitted more applications (aRR 1.32, 95% CI 1.10-1.58) and requested more self-tests (aRR 1.40, 95% CI 1.22-1.60) than each noninfluencer.

When the analysis was limited to indexes with at least one alter who returned a test, sexual health influencers distributed to more alters without prior HIV testing compared to noninfluencers, but this difference was not significant after controlling for the volunteer status of the index (aRR 1.28, 95 CI% 0.85-1.92). Sexual health influencers also performed more simultaneous testing with their alters, and distributed to more alters with HIV-reactive and syphilis-reactive test results compared to noninfluencers, but these differences were not statistically significant (Multimedia Appendix 3 and Multimedia Appendix 4).

During analysis, one outlier sexual health influencer was found to have distributed to 37 alters. To account for potential skew, data from this outlier were excluded in repeat analysis. Differences between sexual health influencers and noninfluencers in the self-test distribution remained the same after excluding this “superdistributer” outlier (see Multimedia Appendix 5 and Multimedia Appendix 6).

Comparing the Alters of Sexual Health Influencers and Noninfluencers

Alters of sexual health influencers versus noninfluencers had some key differences. More alters of sexual health influencers were registered with a rural residence (45.5% vs 23.8%, *P*<.001) and had below college-level education (57.7% vs 37.1%, *P*<.001) compared to alters of noninfluencers. Additionally, more alters of sexual health influencers had more than one casual male partner in the past 6 months (25.2% vs 11.9%, *P*=.005) compared to alters of noninfluencers. Detailed characteristics of the alter participants are provided in Table 2.

Table 2. Characteristics of responding alter participants in Zhuhai, China, 2018-2019.

Characteristic	All alters (N=266)	Alters of sexual health influencers (n=123)	Alters of noninfluencers (n=143)	*P* value
Age (years), mean, SD	29 (7.0)	28 (6.6)	30 (7.3)	.08
Sex at birth, n (%)				.13
Male	262 (98.5)	123 (100.0)	139 (97.2)	
Female	4 (1.5)	0 (0.0)	4 (2.8)	
Residence, n (%)				<.001
Rural	90 (33.8)	56 (45.5)	34 (23.8)	
Urban	176 (66.2)	67 (54.5)	109 (76.2)	
Education, n (%)				<.001
Less than college	124 (46.6)	71 (57.7)	53 (37.1)	
College or more	142 (53.4)	52 (42.3)	90 (62.9)	
Sexual orientation, n (%)				.87
Gay	176 (68.2)	82 (66.7)	94 (65.7)	
Bisexual/other	90 (33.8)	41 (33.3)	49 (34.3)	
>1 male casual partner in past 6 months, n (%)				.005
Yes	48 (18.1)	31 (25.2)	17 (11.9)	
No	218 (82.0)	92 (74.8)	126 (88.1)	

Discussion

Principal Findings

Our implementation study demonstrated that in China, being an MSM sexual health influencer was associated with encouraging more alters to self-test for HIV/syphilis than being a noninfluencer. More importantly, sexual health influencers were associated with alters from rural regions of China and with less education, factors associated with limited access to HIV testing [21,22]. Alters of sexual health influencers also sought more casual sexual partners and thus were at greater risk of HIV infection [23]. These findings suggest that sexual health influencers and noninfluencers can reach different groups of individuals for testing. Our study extends the existing literature on HIV/syphilis self-test distribution by examining the role of sexual health influence. Our findings can inform future studies to increase the reach of self-tests for MSM and other populations at risk of HIV.

We found that being a sexual health influencer was associated with encouraging more alters to self-test than being a noninfluencer. This is consistent with prior findings that MSM peer leaders selected for their social influence were more effective at increasing HIV testing than nonpeer-led interventions [6-8,15,24]. Moreover, sexual health influencers applied for more self-tests than noninfluencers, suggesting greater engagement in self-test promotion. Although being a sexual health influencer was associated with encouraging more first-time testers to test, this was confounded by CBO volunteer status, suggesting that being a volunteer also affected self-test distribution capacity. CBO volunteers have experience in peer engagement even if they do not have other socially influential traits. Our results indicate that the self-administered sexual health influencer scale [17] can identify influential MSM to expand the reach of HIV and syphilis self-tests, and training can help them reach subgroups at higher risk such as MSM without prior testing.

Being a sexual health influencer was also associated with reaching more alters with lower education, from a rural residence, and with more casual sexual partners. These findings suggest that sexual health influencers could better reach individuals at greater risk of HIV but with less access to health care. Several studies have found alters to reach more MSM at high risk of HIV/STIs with low access to testing compared to other recruitment strategies [9,25,26]. Highly influential MSM may help reach individuals with the least access to HIV care and who are often missed by random seeding of interventions [3]. Our findings indicate that self-identified sexual health influencers are low cost and may be effective seeds for HIV/syphilis self-test distribution.

In this study, sexual health influencers were more likely to have disclosed their sexual orientation and to be CBO volunteers compared to noninfluencers. This is consistent with prior findings that MSM sexual health influencers were more likely to disclose their sexual orientation and have greater community engagement [17]. Shared experience between influencers and their target audience (ie, sexual orientation) is known to contribute to the effectiveness of interventions [18]. Moreover, volunteer experience and disclosed sexual orientation may contribute to greater social influence. These findings reinforce the correlation between sexual health influence and volunteerism among MSM, which may contribute to self-test distribution capability. The sexual health influencer scale identified many MSM who were not volunteers but nonetheless had influential qualities and were associated with greater self-test distribution capability, suggesting that sexual health influencers can expand an existing pool of volunteers and reach into untapped social networks.

Although we did not collect data on why sexual health influencers are effective at promoting HIV/syphilis testing, we propose some mechanisms based on existing theories underlying the mechanisms of peer-based health interventions. Our sexual health influencer scale selected for MSM who are socially visible and are sought by peers for advice and information. Based on social comparison theory, sexual health influencers may act as models of health and self-improvement in the MSM community, providing an upward comparison to which other MSM aspire [18]. Sexual health influencers are also experienced in providing social support, which suggests that they may be skilled at providing calm and reducing stress [18] while promoting health behaviors that can cause significant distress for the alter MSM. Few studies on network-based sexual health interventions describe theoretical mechanisms for their efficacy [3]. Future studies should elicit mechanisms that demonstrate the effectiveness of sexual health influencers.

Limitations

Our study has several limitations. First, this was a secondary analysis, which identified sexual health influencers retrospectively. We found correlations between sexual health influence and self-test distribution, but other unaccounted variables may explain the different rates of distribution, such as specific characteristics of an index's social network, influential characteristics unrelated to sexual health, and geographic access between the index and alters. Analyses for alters and indexes were also performed retrospectively, and we could not capture alters who received a self-test but did not return their results. This may have resulted in an underestimation of index self-test distribution rates. Nonetheless, the higher rate of test result through sexual health influencer return may indicate better linkage to care compared to that through noninfluencers. Second, participants were recruited online and required access to the internet for follow-up. Thus, our study only captured MSM with internet access and may have limited representativeness of MSM who do not use online social media. Third, our study was implemented through a well-established gay CBO in a populous city using their existing and popular self-testing platform. This lent our study credibility from participants but limits the generalizability of our findings. Our study may exclude MSM who are not part of any MSM network, and may not be applicable to some regions such as rural communities with weaker MSM networks. Finally, the sexual health influencer scale has only been studied among MSM in East Asia to date, and may have limited generalizability to other cultural and linguistic settings.

Implications

Our study has implications for future HIV/syphilis self-test research and implementation. Primarily, we identified influential individuals using a simple, self-report sexual health influencer scale. This is important when considering the resource intensiveness and complexity of most social network interventions [16]. Simplicity allows for greater sustainability, especially where resource constraints limit social network mapping. Self-identification is an established, low-cost strategy for defining preexisting social influence, but is infrequently utilized due to concerns of subjectivity [27,28]. Nonetheless, we found self-reported sexual health influence to correlate with objective advantages. The sexual health influencer scale can be easily adopted by community programs to identify effective health promoters and educators, particularly when introducing a new health behavior to a marginalized community such as MSM. Our findings indicate social influence to be a core component of effective sexual health promotional campaigns, which should be adopted in sexual health promotion policies. Future randomized controlled studies should test the sexual health influencer scale as an intervention to promote HIV self-testing.

Conclusions

Our study indicates that sexual health influencers are important for encouraging social network–based HIV/syphilis testing. Our findings are notable for the greater distribution by sexual health influencers compared to noninfluencers and increasing access to MSM linked to rural regions, where gay venues and health care facilities may be less accessible, and to MSM with lower education and at higher risk of HIV. Our study can inform future implementation research on social network targeting for HIV self-testing and sexual health interventions.

Acknowledgments

The authors thank the Xutong volunteers, Zhuhai Center for Disease Control research staff, and SESH research assistants for their help and support in implementing this study, as well as all participants for their cooperation. This work was supported by the National Key Research and Development Program of China (2017YFE0103800), Academy of Medical Sciences and the Newton Fund (NIF\R1\181020), National Institutes of Health (NIAID 1R01AI114310-01, NIAID K24AI143471), NIMH (R34MH109359 and R34MH119963), National Science and Technology Major Project (2018ZX10101-001-001-003), National Nature Science Foundation of China (81903371 and 81703282), Guangdong Medical Science and Technology Research Fund (A2020509), Zhuhai Medical and Health Science and Technology Plan Project (20181117A010064), and Shenzhen Philosophy and Social Sciences Fund (SZ2020C026). The funders had no role in any process of this study.

Authors' Contributions

WT and DW conceived the idea and led the project. NY and DW cleaned and analyzed the data. NY and YZ wrote the first draft of this manuscript. SH, XH, XL, XJ, YW, WH, HB, and WD implemented the project and collected data. Other authors provided constructive comments and edited the manuscript. All authors approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The six-item sexual health influencer scale.
[DOCX File , 13 KB - jmir_v23i6e24303_app1.docx]

Multimedia Appendix 2
Distribution of sexual health influence score and distribution statistics for index participants.
[DOCX File , 102 KB - jmir_v23i6e24303_app2.docx]

Multimedia Appendix 3
Self-test application and distribution by sexual health influencer and noninfluencer index men who have sex with men in Zhuhai, China, 2018-2019 (N=371).
[DOCX File , 14 KB - jmir_v23i6e24303_app3.docx]

Multimedia Appendix 4
Self-testing behaviors and outcomes of alters reached by sexual health influencer and noninfluencer index men who have sex with men in Zhuhai, China, 2018-2019 (N=116). Data include only indexes with at least one alter who returned a test result.
[DOCX File , 14 KB - jmir_v23i6e24303_app4.docx]

Multimedia Appendix 5
Self-test application and distribution by sexual health influencer and noninfluencer index men who have sex with men in Zhuhai, China, 2018-2019 excluding 1 superdistributor (outlier) influencer (N=370).
[DOCX File , 14 KB - jmir_v23i6e24303_app5.docx]

Multimedia Appendix 6
Self-testing behaviors and outcomes of alters reached by sexual health influencer and noninfluencer index men who have sex with men in Zhuhai, China, 2018-2019 excluding 1 superdistributor (outlier) influencer (N=115). Data include only indexes having at least one alter who returned a verified test result.
[DOCX File , 14 KB - jmir_v23i6e24303_app6.docx]

References

1. Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. Science 2009 Feb 13;323(5916):892-895. [doi: 10.1126/science.1165821] [Medline: 19213908]
2. Wang K, Brown K, Shen S, Tucker J. Social network-based interventions to promote condom use: a systematic review. AIDS Behav 2011 Oct;15(7):1298-1308 [FREE Full text] [doi: 10.1007/s10461-011-0020-1] [Medline: 21811843]
3. Harling G, Tsai AC. Using social networks to understand and overcome implementation barriers in the global HIV response. J Acquir Immune Defic Syndr 2019 Dec;82(Suppl 3):S244-S252 [FREE Full text] [doi: 10.1097/QAI.0000000000002203] [Medline: 31764260]
4. Kelly JA, Murphy DA, Sikkema KJ, McAuliffe TL, Roffman RA, Solomon LJ, et al. Randomised, controlled, community-level HIV-prevention intervention for sexual-risk behaviour among homosexual men in US cities. Community HIV Prevention Research Collaborative. Lancet 1997 Nov 22;350(9090):1500-1505. [doi: 10.1016/s0140-6736(97)07439-4] [Medline: 9388397]
5. Amirkhanian YA, Kelly JA, Kabakchieva E, Kirsanova AV, Vassileva S, Takacs J, et al. A randomized social network HIV prevention trial with young men who have sex with men in Russia and Bulgaria. AIDS 2005 Nov 04;19(16):1897-1905. [doi: 10.1097/01.aids.0000189867.74806.fb] [Medline: 16227798]
6. Ko N, Hsieh C, Wang M, Lee C, Chen C, Chung A, et al. Effects of Internet popular opinion leaders (iPOL) among Internet-using men who have sex with men. J Med Internet Res 2013 Feb 25;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]

7. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015 Jan;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
8. Young SD, Holloway I, Jaganath D, Rice E, Westmoreland D, Coates T. Project HOPE: online social network changes in an HIV prevention randomized controlled trial for African American and Latino men who have sex with men. Am J Public Health 2014 Sep;104(9):1707-1712. [doi: 10.2105/ajph.2014.301992]
9. Lightfoot MA, Campbell CK, Moss N, Treves-Kagan S, Agnew E, Kang Dufour MS, et al. Using a social network strategy to distribute HIV self-test kits to African American and Latino MSM. J Acquir Immune Defic Syndr 2018 Sep 01;79(1):38-45. [doi: 10.1097/QAI.0000000000001726] [Medline: 29771792]
10. Lippman SA, Lane T, Rabede O, Gilmore H, Chen YH, Mlotshwa N, et al. High acceptability and increased HIV-testing frequency after introduction of HIV self-testing and network distribution among South African MSM. J Acquir Immune Defic Syndr 2018 Mar 01;77(3):279-287 [FREE Full text] [doi: 10.1097/QAI.0000000000001601] [Medline: 29210826]
11. Masters SH, Agot K, Obonyo B, Napierala Mavedzenge S, Maman S, Thirumurthy H. Promoting partner testing and couples testing through secondary distribution of HIV self-tests: a randomized clinical trial. PLoS Med 2016 Nov;13(11):e1002166 [FREE Full text] [doi: 10.1371/journal.pmed.1002166] [Medline: 27824882]
12. Thirumurthy H, Masters SH, Mavedzenge SN, Maman S, Omanga E, Agot K. Promoting male partner HIV testing and safer sexual decision making through secondary distribution of self-tests by HIV-negative female sex workers and women receiving antenatal and post-partum care in Kenya: a cohort study. Lancet HIV 2016 Jun;3(6):e266-e274 [FREE Full text] [doi: 10.1016/S2352-3018(16)00041-2] [Medline: 27240789]
13. Choko AT, Nanfuka M, Birungi J, Taasi G, Kisembo P, Helleringer S. A pilot trial of the peer-based distribution of HIV self-test kits among fishermen in Bulisa, Uganda. PLoS One 2018 Nov 29;13(11):e0208191 [FREE Full text] [doi: 10.1371/journal.pone.0208191] [Medline: 30496260]
14. Gichangi A, Wambua J, Mutwiwa S, Njogu R, Bazant E, Wamicwe J, et al. Impact of HIV self-test distribution to male partners of ANC clients: results of a randomized controlled trial in Kenya. J Acquir Immune Defic Syndr 2018 Dec 01;79(4):467-473 [FREE Full text] [doi: 10.1097/QAI.0000000000001838] [Medline: 30148731]
15. Campbell CK, Lippman SA, Moss N, Lightfoot M. Strategies to increase HIV testing among MSM: a synthesis of the literature. AIDS Behav 2018 Aug 17;22(8):2387-2412. [doi: 10.1007/s10461-018-2083-8] [Medline: 29550941]
16. Shelton RC, Lee M, Brotzman LE, Crookes DM, Jandorf L, Erwin D, et al. Use of social network analysis in the development, dissemination, implementation, and sustainability of health behavior interventions for adults: A systematic review. Soc Sci Med 2019 Jan;220:81-101 [FREE Full text] [doi: 10.1016/j.socscimed.2018.10.013] [Medline: 30412922]
17. Wu D, Tang W, Lu H, Zhang TP, Cao B, Ong JJ, et al. Leading by example: web-based sexual health influencers among men who have sex with men have higher HIV and syphilis testing rates in China. J Med Internet Res 2019 Jan 21;21(1):e10171 [FREE Full text] [doi: 10.2196/10171] [Medline: 30664490]
18. Simoni JM, Franks JC, Lehavot K, Yard SS. Peer interventions to promote health: conceptual considerations. Am J Orthopsychiatry 2011 Jul;81(3):351-359 [FREE Full text] [doi: 10.1111/j.1939-0025.2011.01103.x] [Medline: 21729015]
19. Kelly JA. Popular opinion leaders and HIV prevention peer education: resolving discrepant findings, and implications for the development of effective community programmes. AIDS Care 2004 Feb;16(2):139-150. [doi: 10.1080/09540120410001640986] [Medline: 14676020]
20. Wu D, Zhou Y, Yang N, Huang S, He X, Tucker J, et al. Social media-based secondary distribution of HIV/syphilis self-testing among Chinese men who have sex with men. Clin Infect Dis 2020 Jun 26:ciaa825. [doi: 10.1093/cid/ciaa825] [Medline: 32588883]
21. Li X, Fang X, Lin D, Mao R, Wang J, Cottrell L, et al. HIV/STD risk behaviors and perceptions among rural-to-urban migrants in China. AIDS Educ Prev 2004 Dec;16(6):538-556 [FREE Full text] [doi: 10.1521/aeap.16.6.538.53787] [Medline: 15585430]
22. Li X, Lu H, Ma X, Sun Y, He X, Li C, et al. HIV/AIDS-related stigmatizing and discriminatory attitudes and recent HIV testing among men who have sex with men in Beijing. AIDS Behav 2012 Apr 16;16(3):499-507 [FREE Full text] [doi: 10.1007/s10461-012-0161-x] [Medline: 22350831]
23. Feng Y, Wu Z, Detels R, Qin G, Liu L, Wang X, et al. HIV/STD prevalence among men who have sex with men in Chengdu, China and associated risk factors for HIV infection. J Acquir Immune Defic Syndr 2010 Feb;53(Suppl 1):S74-S80 [FREE Full text] [doi: 10.1097/QAI.0b013e3181c7dd16] [Medline: 20104114]
24. Okoboi S, Lazarus O, Castelnuovo B, Nanfuka M, Kambugu A, Mujugira A, et al. Peer distribution of HIV self-test kits to men who have sex with men to identify undiagnosed HIV infection in Uganda: A pilot study. PLoS One 2020 Jan 23;15(1):e0227741 [FREE Full text] [doi: 10.1371/journal.pone.0227741] [Medline: 31971991]
25. Tun W, Vu L, Dirisu O, Sekoni A, Shoyemi E, Njab J, et al. Uptake of HIV self-testing and linkage to treatment among men who have sex with men (MSM) in Nigeria: A pilot programme using key opinion leaders to reach MSM. J Int AIDS Soc 2018 Jul 22;21(Suppl 5):e25124 [FREE Full text] [doi: 10.1002/jia2.25124] [Medline: 30033680]
26. Das A, George B, Ranebennur V, Parthasarathy MR, Shreenivas GS, Todankar P, et al. Getting to the first 90: incentivized peer mobilizers promote HIV testing services to men who have sex with men using social media in Mumbai, India. Glob Health Sci Pract 2019 Sep 26;7(3):469-477. [doi: 10.9745/ghsp-d-19-00094]

27. Valente TW, Pumpuang P. Identifying opinion leaders to promote behavior change. Health Educ Behav 2007 Dec 21;34(6):881-896. [doi: 10.1177/1090198106297855] [Medline: 17602096]
28. Aral S, Muchnik L, Sundararajan A. Engineering social contagions: Optimal network seeding in the presence of homophily. Netw Sci 2013 Jul 30;1(2):125-153. [doi: 10.1017/nws.2013.6]

Abbreviations

aRR: adjusted rate ratio
CBO: community-based organization
MSM: men who have sex with men
POL: popular opinion leader
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 14.09.20; peer-reviewed by P Martin, J Xu; comments to author 10.11.20; revised version received 02.12.20; accepted 14.04.21; published 01.06.21.

Please cite as:
Yang N, Wu D, Zhou Y, Huang S, He X, Tucker J, Li X, Smith KM, Jiang X, Wang Y, Huang W, Fu H, Bao H, Jiang H, Dai W, Tang W
Sexual Health Influencer Distribution of HIV/Syphilis Self-Tests Among Men Who Have Sex With Men in China: Secondary Analysis to Inform Community-Based Interventions
J Med Internet Res 2021;23(6):e24303
URL: https://www.jmir.org/2021/6/e24303
doi:10.2196/24303
PMID:34061035

Original Paper

Patients and Professionals as Partners in Hypertension Care: Qualitative Substudy of a Randomized Controlled Trial Using an Interactive Web-Based System Via Mobile Phone

Ulrika Andersson[1], MD; Ulrika Bengtsson[2,3], PhD; Agneta Ranerup[4], PhD; Patrik Midlöv[1], MD, PhD; Karin Kjellgren[3,5], PhD

[1]Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden
[2]Institute of Health and Care Sciences, University of Gothenburg, Gothenburg, Sweden
[3]University of Gothenburg Centre for Person-Centred Care, University of Gothenburg, Gothenburg, Sweden
[4]Department of Applied Information Technology, University of Gothenburg, Gothenburg, Sweden
[5]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Ulrika Andersson, MD
Department of Clinical Sciences Malmö
Lund University
PO Box 50332
Malmö, 202 13
Sweden
Phone: 46 070 6251729
Email: ulrika.andersson.4418@med.lu.se

Abstract

Background: The use of technology has the potential to support the patient´s active participation regarding treatment of hypertension. This might lead to changes in the roles of the patient and health care professional and affect the partnership between them.

Objective: The aim of this qualitative study was to explore the partnership between patients and health care professionals and the roles of patients and professionals in hypertension management when using an interactive web-based system for self-management of hypertension via the patient's own mobile phone.

Methods: Focus group interviews were conducted with 22 patients and 15 professionals participating in a randomized controlled trial in Sweden aimed at lowering blood pressure (BP) using an interactive web-based system via mobile phones. The interviews were audiorecorded and transcribed and analyzed using thematic analysis.

Results: Three themes were identified: the technology, the patient, and the professional. The technology enabled documentation of BP treatment, mainly for sharing knowledge between the patient and the professional. The patients gained increased knowledge of BP values and their relation to daily activities and treatment. They were able to narrate about their BP treatment and take a greater responsibility, inspired by new insights and motivation for lifestyle changes. Based on the patient's understanding of hypertension, professionals could use the system as an educational tool and some found new ways of communicating BP treatment with patients. Some reservations were raised about using the system, that it might be too time-consuming to function in clinical practice and that too much measuring could result in stress for the patient and an increased workload for the professionals. In addition, not all professionals and patients had adopted the instructions regarding the use of the system, resulting in less realization of its potential.

Conclusions: The use of the system led to the patients taking on a more active role in their BP treatment, becoming more of an expert of their BP. When using the system as intended, the professionals experienced it as a useful resource for communication regarding BP and lifestyle. Patients and professionals described a consultation on more equal grounds. The use of technology in hypertension management can promote a constructive and person-centered partnership between patient and professional. However, implementation of a new way of working should bring benefits and not be considered a burden for the professionals. To establish a successful partnership, both the patient and the professional need to be motivated toward a new way of working.

Trial Registration: ClinicalTrials.gov NCT03554382; https://clinicaltrials.gov/ct2/show/NCT03554382

(J Med Internet Res 2021;23(6):e26143) doi:10.2196/26143

KEYWORDS

eHealth; digital health; hypertension; mobile phones; patient-professional partnership; person-centered care; self-management

Introduction

Background

Medical advances and better living conditions have led to increasing lifespans and a growing population living with chronic conditions such as hypertension [1]. With limited health care resources, new, more effective ways of managing chronic conditions need to be developed [2]. Patients cannot be regarded as passive recipients of care but will need to be the main providers of care for themselves. With this, the role of health care professionals will also need to change from being the expert provider of care to being a cocreator of care with the patient [3,4]. During 2020 and the COVID-19 pandemic, the need for this has become even more evident. Patients need to be able to perform effective self-management in their homes and not be dependent on visiting or using health care facilities [5]. However, self-managing high blood pressure (BP) is something patients do every day by choosing what to eat, deciding whether to exercise, trying to decrease stress, and remembering to take their prescribed medication [3]. Health care professionals have an important role to play in supporting patients to self-manage, ideally working in partnership with patients [6].

A European standard for a minimum level of patient involvement was recently established with the aim to support a wider implementation of person-centered care (PCC) [7]. PCC is a health care approach where the patient's subjective perception of illness and their preferences and values are the starting point for the care process. Partnership between patient and professional, as well as patient narratives and shared documentation, are considered key concepts in PCC. Within the narrative and examination, the patient's need of care, prerequisites, resources, and obstacles are identified and documented together with the patient [8]. Attributes defining partnership vary in different publications, but shared decision making, shared knowledge, communication, and shared power are commonly mentioned. The consequences are described as empowerment of the patient and improved health outcome and health care utilization [9-11]. Patients appear to value other aspects of partnership than formal frames, appreciating proximity and receptive communication more than shared documentation and goal setting [12]. Using technological tools in health care may strengthen the potential for patient self-management, and the understanding and practice of partnership between patient and professional might change as a result [13].

Objectives

Using an interactive information technology system requires interaction between patients and professionals, thus possibly affecting the patient-professional partnership. New roles for patients and professionals may be enabled. To date, there is limited research on how using technological tools in BP treatment affects the relationship between the patient and health care professional.

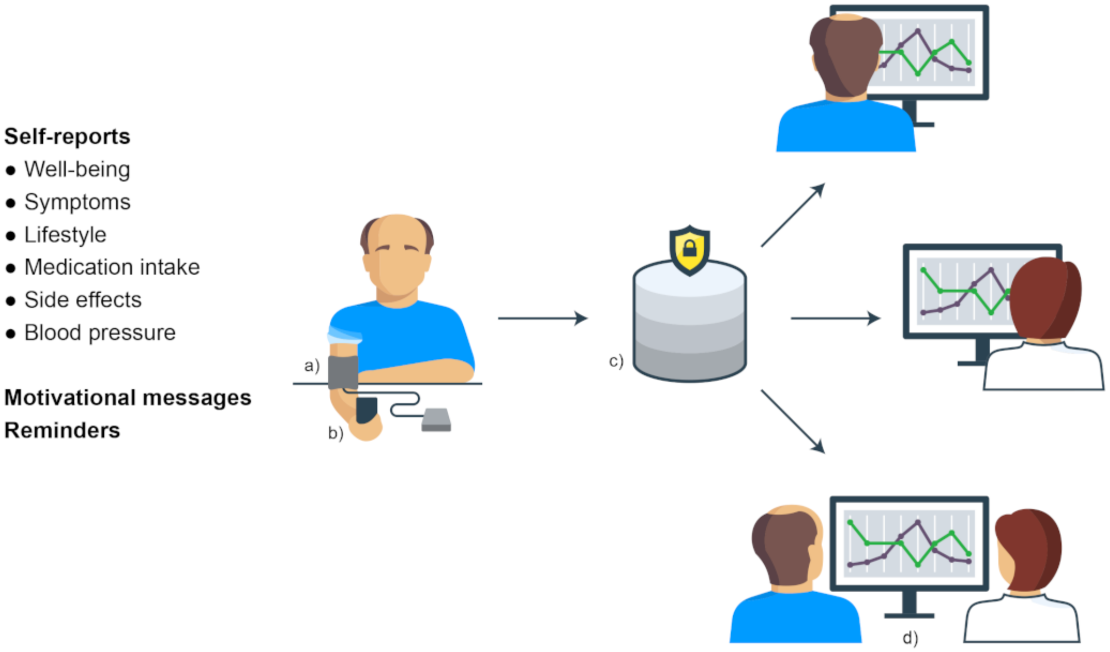
The objectives of this study were to explore the partnership between patients and health care professionals and further the roles of patient and professional when using an interactive web-based system for self-management of hypertension via the patient's own mobile phone.

Methods

Study Design

This study builds on a previously described interactive web-based communication system for self-management of hypertension called CQ (developed by Circadian Questions AB and referred to in this paper as “the system”). The system has been described in earlier publications [14-16], and an overview can be seen in Figure 1. During the planning, execution, and evaluation of the components of the system in the pilot project, the participating patients and professionals were actively involved [14-18]. The system was found to be relevant and easy to use [19], resulting in a significantly decreased BP for the participants (systolic BP –7 mm Hg and diastolic BP –4.9 mm Hg) [20]. Furthermore, use of the system was considered a resource for PCC and a more autonomous, knowledgeable, and active patient [21,22].

Figure 1. Overview of the interactive web-based communication system: (a) blood pressure device; (b) self-reports, reminders, and optional motivational messages via patient's own mobile phone; (c) database where self-reports are saved; and (d) secure web portal available to patients and professionals for data visualization.

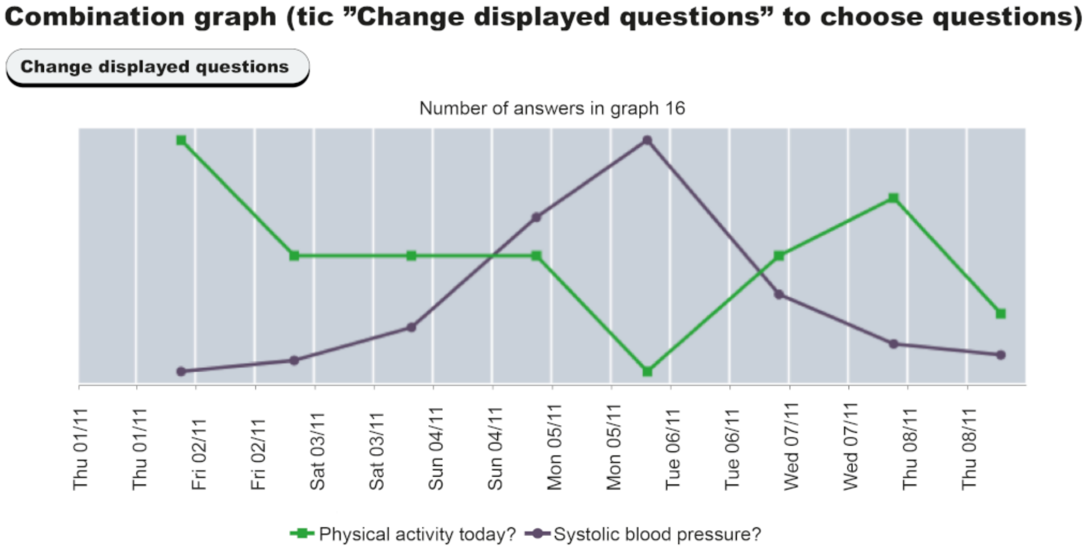


The system described in Figure 1 is now being tested in a randomized controlled trial (RCT; Person-Centeredness in Hypertension Management Using Information Technology [PERHIT]), including 900 patients with hypertension equally allocated to an intervention and a control group. The trial is conducted in primary care in 4 health care regions in southern Sweden. The aim of the trial is to lower BP in patients with hypertension in primary care. In addition, person-centeredness, patient self-reports such as daily life activities, and awareness of risk will be evaluated [23].

In short, the intervention consists of the following:

- Start-up meeting was scheduled with a nurse or physician at the local primary health care center (PHCC) where instructions were given about how to use the system at home, including measuring BP daily. Questions regarding side effects were selected according to the patients' medication. Patients could choose to receive different relevant motivational messages on different days of the week. The messages were in the form of motivational questions and were intended to function as an inspiration for healthy choices (eg "Nice walk at lunch today?"). Patients also received a manual of the system and were advised to watch videos on BP measurement and how to enter data via their mobile phones.
- During 8 consecutive weeks, patients used the system at home and reported BP, symptoms, medication intake, side effects, lifestyle, and well-being. After log-in, patients and professionals had access to visualization of self-reported data in graphs via a secure web portal. All data was saved in a secure database, not in the mobile phones.
- Follow-up consultation was scheduled with a nurse or physician at the local PHCC after finishing the 8-week intervention. Professionals were encouraged to discuss graphs with patients. An example of a system graph is presented in Figure 2.
- Follow-up consultation was scheduled for 12 months after trial began.

Figure 2. Graph showing correlation of physical activity with blood pressure, as shown to participants.



Several interventions comprising mHealth (the use of mobile devices in health care) and hypertension have shown promising results [24-26]. However, the evidence is scarce, and several research studies have called for large RCTs with mHealth interventions that involve more patients for a longer time period [27-30].

In this qualitative study, we conducted focus group interviews with patients and professionals participating in PERHIT. The Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist was used to ensure rigor in reporting the study design and conduction [31].

Recruitment and Participants

Four PHCCs participating in PERHIT in different geographical and socioeconomic areas were strategically selected to reflect a broad socioeconomic area. Two of the PHCCs were located in midsize cities, one in a larger city suburb, and one in a smaller city. Patients and professionals were contacted by the staff at the PHCC to take part in focus group interviews.

At the time of the interviews, all patients had completed their 8-week intervention and attended the follow-up consultation with their nurse or physician. The time elapsed from the completion of the intervention to the interview varied between the patients from 1 week to 3 months (median 31 days). The inclusion criteria for the patients were the same as for the PERHIT study: aged older than 18 years, diagnosis of hypertension, treatment with at least one antihypertensive drug, and understanding of Swedish in order to be able to provide informed consent and make use of the system using the mobile phone for answering questions [23]. The inclusion criteria for the professionals were being a nurse or physician at the PHCC and having experience working with the PERHIT study.

Since only 2 to 4 professionals were involved in the study at each site, other professionals in the PERHIT study from nearby sites were also approached and asked to participate in the same interview. In total, professionals from 8 different PHCCs contributed to the study.

Data Generation

Prior to the focus group interviews, 2 semistructured interview guides were developed by the research team, one for the patient groups and one for the health care professional groups. A test interview with mock patients was conducted prior to the first interview to evaluate the questions, resulting in some changes to the interview guide. After the first focus group interview with patients, it was obvious that a few questions needed to be rephrased. These were minor changes, and the material from the interview was still considered useful. No changes to the interview guides were made after that. Interview topics are presented in Textbox 1. Interviews began after introductions, small talk, and reiteration of the research goal [32].

Textbox 1. Interview topic and subtopic list.

Hypertension and support:

- Current blood pressure treatment, drugs/lifestyle (patients)
- Perceptions of the most important treatment of blood pressure (professionals)
- Support in blood pressure treatment for patients from primary health care center in usual care

Patient-centered care and partnership:

- Perceptions of patient-centered care
- Perceptions of partnership or how to collaborate with health care professionals/patients, generally and specifically during the intervention
- Experiences from the follow-up consultation after 8 weeks
- Experiences of discussing patients in need of support in blood pressure treatment
- Perceptions of patients' role in blood pressure treatment

Using the system and technology:

- Experiences of using the technology and how it was used during the 8-week intervention
- Perceptions of motivational messages
- How/if using the system has affected everyday life (patients)
- How/if using the system has affected working methods in blood pressure treatment (professionals)
- Experiences of using other technical systems for chronic disease in health care

Focus group interviews were held at the PHCCs from June 2019 to January 2020. A total of 22 patients participated in 4 focus group interviews, with 4 to 7 patients in each. Three focus group interviews, with 4 to 6 professionals each (n=15 total), were also conducted. No compensation was offered to the participants except for coffee and fruit during the interview. The duration of focus group interviews varied from 64 to 97 minutes and were held in Swedish. UA (first author) was the moderator of the focus group interviews. UB (second author), who is experienced in qualitative research, assisted and took notes. Prior to the focus group interview, UA had been in contact with the patients and professionals by telephone or email to set a date and time for the interview. No other relationship prior to the interview was established. At the interviews, researchers presented themselves with their occupation and as members of the research group conducting the RCT. Only the participants and researchers were present at the interviews.

Data Analysis

The interviews were audiorecorded and transcribed verbatim. They were also videorecorded, with the purpose to serve as an aid for memory during the analysis phase. Thematic analysis according to Braun and Clarke [33] was used on the dataset, since it is a flexible method when performing qualitative analysis, allowing for both an inductive and deductive approach to the data [34].

The recordings of the interviews were listened through several times and the anonymized transcripts were checked against the recordings for errors by the first author (UA). During this phase, initial thoughts and ideas were noted. UA, UB, and KK (last author) read the transcripts repeatedly. UA created initial codes by systematically going through all the interview transcripts without a predefined coding frame. Interviews with patients and professionals were coded simultaneously using NVivo software (version 12, QSR International). The initial codes were compared and organized into common categories, which were discussed by UA, UB, and KK. Since we were interested in a specific aspect of the participants' experiences—how using the system affected the experience of partnership between patients and professionals—we then used a deductive approach, inspired by previous research concerning the concept of partnership [8-11] and partnership and technology [13,35]. The initial codes were reviewed and arranged in preliminary themes and subthemes, focused on aspects of partnership. A narrative description of the preliminary themes and a thematic map were created and discussed by UA and KK. Themes were reviewed and checked against the datasets. In the process of defining and naming the themes, UA, KK, and AR collaborated and discussed until consensus was reached. A detailed description of each theme was developed, and informative names for the themes were established. To further visualize the themes, descriptive excerpts were identified.

Ethical Considerations

The study was approved by the regional ethical review board in Lund (2017/311 and 2019/00036). Participants were given oral and written information about the study before they signed a consent form. All transcripts were anonymized to ensure confidentiality. The study was registered with ClinicalTrials.gov [NCT03554382].

Results

Study Sample

Characteristics of participating patients and professionals are presented in Tables 1 and 2.

Table 1. Characteristics of participating patients (n=22).

Characteristic	Value
Female, n (%)	8 (36)
Age (years), median (range)	65 (46-81)
Age intervals (years), n (%)	
<50	2 (9)
50-70	15 (68)
>70	5 (23)
Country of birth, Sweden, n (%)	22 (100)
Years with hypertension[a], median (range)	6 (1-39)
Number of hypertension drugs, median (range)	2 (1-4)
Marital status, n (%)	
Married	15 (68)
Unmarried	6 (27)
Widowed	1 (5)
Education level[b], n (%)	
Up to high school	4 (18)
High school	7 (32)
University	9 (41)
Employment status, n (%)	
Employed	10 (45)
Retired	12 (55)

[a] Years with hypertension: missing 1 data point.

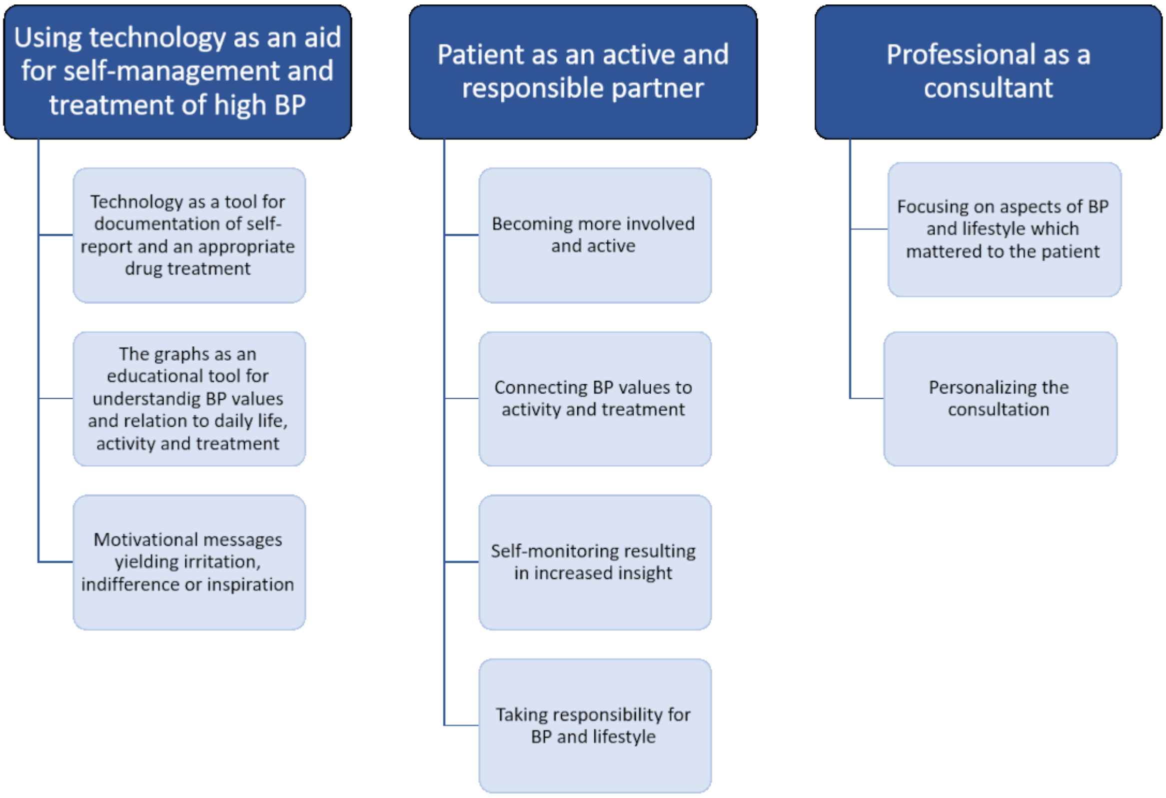
[b] Education level: missing 2 data points.

Table 2. Characteristics of participating health care professionals (n=15).

Characteristic	Value
Female, n (%)	10 (67)
Age (years), median (range)	55 (29-71)
Age intervals (years), n (%)	
<40	3 (20)
40-60	9 (60)
>60	3 (20)
Occupation, n (%)	
Assistant nurse	2 (13)
Nurse	4 (27)
District nurse	3 (20)
Resident physician	1 (7)
General practitioner	5 (33)
Years of experience working with patients with hypertension, median (range)	17 (4-30)

In the analysis of the focus group interviews, 3 actors were identified: the patient, the professional, and the technology. The roles of the different actors are described in 3 themes: using technology as an aid for self-management and treatment of high BP, professional as a consultant, and patient as active and responsible partner. An overview of the themes and subthemes is presented in a thematic map in Figure 3.

Figure 3. Overview of the themes and subthemes.



Using Technology as an Aid for Self-Management and Treatment of High BP

Technology as a Tool for Documentation of Self-Reports and Appropriate Drug Treatment

The professionals considered the different components of the system to be helpful tools in the treatment of high BP. The documentation of the self-reports via the graphs made it possible to communicate more easily about the treatment. If a new drug was prescribed, the patient could follow the effect from day to day, thus becoming more aware of the BP treatment. It was considered a benefit that the patients monitored their BP at home instead of coming to the PHCC. During the intervention, some patients had contacted their nurse or physician when their BP was high, thus acting on high BP values.

Professionals viewed selected patients' graphs during the 8-week intervention if the patient encountered problems adjusting the BP. If the BP was still too high, they contacted the patient and could adjust the drug treatment without the patient having to come to the PHCC. They believed this was educational for them as professionals as well, leading to increased understanding of the variation of BP.

> *And you could go in and see...see when they were running high, if something had happened, so to say, that day. If they were stressed or...if something...and if you saw that they were still running too high, so to say, you called and talked to them and said we need to adjust your medicine.* [Health Care Professional 2 (HCP2)]

For some of the patients, using the system brought a closer and more frequent contact with their prescribing physician. If they had altered their BP medication at the start of or during the intervention, they could with daily measurement report its effect on the BP.

> *But even with close, sort of, contact with M [the patient's physician] which is...it's been short telephone calls where you can...yeah, but he's asked "How are you?" etc. Yeah, this is how it's going now, and then we've been able to change it quickly.* [Patient 13 (P13)]

The professionals' opinions about how feasible it would be to use the system as an integrated part of BP management differed. Not all were positive. Some experienced that it was too time-consuming and did not provide enough benefits to make it worthwhile.

Graphs as an Educational Tool for Understanding BP Values and Relation to Daily Life, Activity, and Treatment

During the follow-up consultation, graphs were used as an educational tool. Through the graphs, the patient could become aware of the normal variability of BP. They could also connect BP variations to physical activity, stress, or medication intake, for example, creating awareness of lifestyle and medication effect on BP. The patients contributed with their explanation of BP variation in relation to their daily activities.

> *But I myself had...in my head I kind of had the idea that now I want to see these graphs for these particular days and what I knew that I...had reported high then and also made a note of it, so to say, and it matched well. Yeah, I thought there was good correlation between these...* [P22]

Not all professionals viewed the graphs with the patients during the follow-up consultations, thus not using the system as intended. In those cases, the professionals expressed that the patients were passive during the follow-up consultation. The BP and lifestyle were not discussed, and instead the patients waited for the professional to introduce the next step in the study procedure. In these cases, the professionals had not adopted the instructions given by the research team regarding the intended use of the system.

Motivational Messages Yielding Irritation, Indifference, or Inspiration

The optional motivational messages included in the system, in the form of motivational questions, were meant to function as an inspiration for healthy choices. Opinions among the patients about the messages differed. Some patients perceived them as irritating since it was not possible to submit an answer. These patients had not been informed (and had not read the manual) about the intention with the messages to function as small reminders not requiring an answer. They thought a positive answer to the questions would generate further information. Others simply ignored the messages, since they disappeared in all the other incoming information in their mobile phones, such as text messages, emails and alerts. Other patients perceived the motivational messages as something positive, finding them an inspiration for healthy choices or considered them a small sign that someone cared about them.

Patient as an Active and Responsible Partner

Becoming More Involved and Active

After using the system, patients were considered by the professionals to be more active in the consultation. They asked questions and wanted to discuss their BP values in relation to the documentation of their daily activities. The professional did not have to lead the conversation as they usually did.

> *Yeah, they were very serious; they had direct questions then, oh yeah, I saw that that day looked like this, what do you say about this, sort of...I didn't have to ask that much; they had their questions for me.* [HCP13]

The patients considered themselves as more involved during the follow-up consultation, since they contributed with their knowledge about how they had felt and their health status. They considered themselves more prepared for the consultations and had thought about questions and what they wanted to discuss with the professional. They also believed this was recognized and confirmed by the professional, who was considered to be interested and attentive.

> *I'd say that you felt more like a participant, because I'd, like, been in this study and knew how I'd felt and how I, like...now it was...I could also offer something and contribute something.* [P6]

Connecting BP Values to Activity and Treatment

Using the system made the patients more aware of how their choices affected their BP and their health, and they reflected upon their days. Being able to measure the BP frequently gave insight into how different BP levels corresponded to daily activities.

> *But, you know, I've noticed right away when I've made that change there, I mean with the exercise and then also training with my dumbbells at home and stuff, that it's had an effect; it has, you know.* [P10]

Not all patients logged in to the website and viewed their reported values in graphs during the intervention. Reasons for not logging in were that they were not aware of the possibility, they were not interested, or they chose to wait until the follow-up consultation. The patients who did view the graphs by themselves thought they were valuable and used them to relate activities or well-being to their BP values. Some patients who were not aware of the possibility to log in to the website kept notes by themselves, writing down the BP and what they had done and in some instances sharing their notes with their nurse or physician. Even if they did not view the graphs, they connected their daily BP value to how they felt or what they had done during the day.

Self-Monitoring Resulted in Increased Insight

By monitoring the BP and relating it to daily life, the patient became the expert on his or her BP. Some patients related that they got to know themselves and their bodies better. By daily monitoring, they could anticipate the BP value when measuring it in the evening. For example, after a stressful day, they expected a high BP value. They became aware of what affected the BP and what they could do about it. Their own responsibility for a successful treatment became clear to them.

> *Yeah, and it...it was a...yeah, it was actually a wake-up call too. That you could do something yourself; that you should do something yourself.* [P6]

Taking Responsibility for BP and Lifestyle

The patients regarded it as their responsibility to contact their physician or nurse when their BP was uncontrolled. They also considered it their responsibility to keep track of their BP and appreciated being able to check their BP at home.

> *So if I felt a little uncomfortable in my body and I went and checked my blood pressure and it was a total disaster, yeah, then I could sound the alarm earlier than if I hadn't had a gauge. So in that way I feel safer now, I think.* [P10]

The professionals related that they saw an increased interest in self-monitoring of BP, even outside the settings of a study. They had noticed that some of their patients had bought BP monitors and used them at home. This was mostly considered positive, although the professionals also thought some patients measured and monitored excessively, which could lead to an increased workload for them and stress for the patient.

The patients considered diet and physical activity important regarding BP treatment. Participation in the study was a motivator for lifestyle changes such as increasing physical activity. They believed they had knowledge about the positive effects of exercise and a healthy diet prior to the study but had not taken it to heart before. Seeing the BP values every day became a reminder and encouragement to do something about the situation. The need to do well and be normal, to have a BP within the target values, was also a motivator for healthier choices. Some patients changed their dietary habits, cutting back on sweets, salt, and licorice. Others had thought about changing their habits but had not yet started.

> *I mean, I think, before I started this study I knew that exercise was the best, but even so...you heard it every time you came down and took your blood pressure and stuff, but yeah...you know, not much happens. But when I started this study I became much more focused; I thought I have to get this blood pressure down—I myself have to help too. So I've actually started exercising much, much more.* [P6]

Professional as a Consultant

Focusing on Aspects of BP and Lifestyle That Mattered to the Patient

The professionals believed using the system contributed to more lifestyle-oriented conversations with the patients. Instead of only focusing on the effect of BP-lowering drugs, they talked about other aspects of high BP, such as how the patient's lifestyle affected BP. The professionals considered the conversation to be more focused on the individual patient's needs and resources than usual BP consultations. The patients said that they could discuss things that were important to them, that either they themselves or the professionals brought up. Some of the professionals expressed that they became more of a consultant for the patient than a lecturing nurse or physician. When the patients were more active during the follow-up consultations, possible lifestyle changes, which were significant for them, came to light and the discussion could focus on that on their terms.

> *And then maybe something turns up...one thing we can help with and work with, but then maybe we can calm down a little with the rest of...because it's this that the patient's a little interested in or feels like I have to...this...I can make a change here, and then we can help with that. Yeah, it was...it made it easier...to have that kind of discussion, I think.* [HCP3]

Personalizing the Consultation

By introducing the system to the patients and looking at the graphs together, the professionals related that they found out more about the patients and learned something new from them. One professional was surprised about how much the knowledge about BP differed between patients; some did not know about the risk of elevated BP or their own target BP. When this came to light, the discussion could be held in a more personalized way.

While using the graph as a visual tool, some of the professionals related that they learned new ways of talking about BP and lifestyle. Despite years of experience of talking to patients about BP, this consultation was considered more rewarding as it was more personal and relevant.

> *In some way I learned to teach people about blood pressure, which I actually hadn't done before. I've seen so many blood pressure patients, but haven't ever had the time to get into this particular person's condition, kind of.* [HCP13]

Thus, not only could the patients gain new knowledge by using the system, the professionals could deepen their understanding of hypertension management.

Partnership

As shown above, the system contributed to several attributes of partnership (see the code list from NVivo in Multimedia Appendix 1). Patients contributed with their knowledge about their health status and situations while professionals contributed with expert knowledge on BP, thus sharing knowledge. The professionals expressed that they learned new things using this working method. Both patients and professionals declared that the consultation was more equal as the patient was more prepared and knowledgeable, thus indicating shared power and shared collaborative decision making.

Discussion

Principal Findings

This study aimed to explore the partnership and roles of patients and professionals in hypertension management when using an interactive web-based system for self-management of hypertension. Focus group interviews with patients and professionals were conducted and analyzed using thematic analysis.

Three themes, on the technology, the patient, and the health care professional, are evident when using the interactive web-based system via mobile phone. The described themes represent one actor each. The system (the technology) is mainly a tool for documentation and sharing knowledge between patient and professional, thus affecting the partnership and how BP is communicated. By using the system, patients gained insight into how BP was affected by their lifestyle and became motivated to make healthier choices. As experts of their BP, they came well prepared to the follow-up consultation and were then able to take on a more active role. The professionals took a more secondary role during the follow-up consultation, controlling the conversation to a lesser extent. They were no longer the only holders of data and knowledge but instead became consultants and support to the patients, contributing with expert knowledge adjusted for the patients' needs. Both patients and professionals described a consultation on more equal terms than usual, thus creating a base for a successful partnership. This was the case described by most of the participating professionals and patients but not shared by all.

Comparison With Prior Work

Previous research has found that self-monitoring of BP enables activation of patients and motivates them to engage in lifestyle changes, favoring self-management [6,36]. By self-monitoring, the patient can provide the data that was previously produced by the health care professional at the clinical encounter. According to Shahaj et al [6], this might potentially challenge the dynamics between patient and professional, which is in line with the findings in this paper. Most of the patients in this study considered it their responsibility to check their BP regularly and adhere to the prescribed treatment. This was mainly viewed as something positive, but it could have potentially negative effects. If the patient using the system is not able to take on the responsibility, for example, not being able to interpret the BP values and acting on high values, the use of the system could be a burden. The system is intended to be used as a complement to the physical meeting and examination in usual care, and thus a patient not being able to use or interpret the system should not receive inferior care compared to treatment as normal. On the other hand, if the patient is able to take on the responsibility and self-manage effectively, the need for physical check-ups is diminished and contact with the health care professional can be managed over the phone or digitally in an effective way.

Wildevuur et al [13] studied how the partnership between patients and professionals is affected by the use of information and communication technology. They found that using information and communication technology in disease management requires an adjustment of the partnership through strengthened potential for self-management and shared analyzing of data. The health care system can be reorganized with new care pathways, where the data provided by the patient can serve as an initiative for treatment. Ultimately, it is the patients' trust in technology and ability to self-manage that shapes the partnership with the professionals, provided that the professionals can adapt to the different needs of different patients.

In our interviews, opinions on using the system differed among professionals. Most of them found the system to be a helpful tool regarding hypertension management, inspiring new ways to talk about hypertension and working with the patient as an equal partner. Others were apprehensive about using it in clinical work since they found it too time-consuming. The professionals' views about the role of technological tools in clinical work also differed; some did not believe it would bring any positive effects while others considered it an inevitable and possibly favorable part of their future working methods. A precondition for technology to enable effective PCC is, according to Wildevuur et al [35], that the technical solution is efficient for both patients and professionals and reduces the pressure on health care systems. In our study, the intervention technology is not integrated in the established health care technology, thus requiring the professionals to work in parallel systems, and this might cause problems. During the interviews, we found that the system was not used as intended in some instances despite a thorough introduction and a user manual. Some of the patients were not aware of the possibility of logging in to the web portal and viewing their reported values in graphs. This opportunity for visual feedback and insights of connections between BP and reported factors was therefore lost. Some of the professionals had not viewed the graphs together with the patients at the follow-up consultation after the 8-week intervention, thus disregarding a large part of the potential use of the system and an important kick-off for lifestyle changes during the rest of the 12 months. A lesson learned is that when introducing a new technical system, the professionals' opinions and preferences about technology need to be acknowledged and considered. The professionals need to receive sufficient education on how to make use of the system in an optimal way and correctly instruct the patients on how to use it and what the benefits are for the patients in doing so. Implementation of a new way of working should bring benefits and not be considered a burden for the professionals. To establish a successful partnership, both the patient and professional need to be motivated about the new way of working.

The optional motivational messages in our study were received with mixed emotions by the patients. Previous interventions, which focused on text message–based lifestyle advice with the aim to lower BP, had shown small or insignificant positive results, indicating that motivational messages might be a part of a successful lifestyle intervention but are not sufficient on their own [37,38]. The irritation some of the patients described about the messages could be attributed to a lack of knowledge of the intention with the messages, highlighting the need for sufficient education of the health care professionals when conducting a study like this. That some patients were aware of the intention of the messages but chose to ignore them indicates that for a lifestyle intervention to be successful, some response or action is necessary. Otherwise, the message will disappear in the amount of information received daily.

Previous research has shown that follow-up consultations regarding hypertension are usually dominated by the professional and mainly focused on effect of drug treatment on BP [39]. As a contrast to this, during the follow-up consultations in this study, the focus was more on lifestyle and its relation to BP. The visualization of BP and lifestyle in graphs was considered valuable and contributed to the change of focus. The patients described that with insight gained by using the system at home and during the consultation came motivation to make lifestyle changes. This can lead to improved physical and psychosocial well-being beyond the effect on BP levels.

When used as intended, the system was found to be a resource for a person-centered approach in hypertension management. After implementing the system for 8 weeks, the patients could express their views and experiences of high BP. Both patients and professionals could contribute with knowledge during the follow-up consultation. The graphs could serve as documentation, shared by the patient and the professional.

To further analyze the potential benefit of using this system, future studies could focus on testing in other clinical or cultural settings such as hospital clinics with outpatient care or in other countries with a more diverse population. As reported by Samkange-Zeeb et al [40], it is important to consider migration background and language competency to make information and services via the internet accessible in diverse groups.

Strengths and Limitations of the Study

A strength of this study is that it builds on previous work and confirms results found in the pilot project regarding the potential of the system to support patient self-management. By including experiences from both patients and professionals in the study, a more comprehensive dataset was obtained. Both perspectives must be identified before implementation in clinical practice.

This study has some limitations. During recruitment for the focus group interviews, we aimed to put together groups with a diversity of men and women from different socioeconomic and cultural backgrounds and in different age groups. We therefore approached PHCCs in different socioeconomic areas. The number of eligible patients, with the diversity above, per PHCC was limited and not all proposed patients agreed to participate. We therefore had to make some compromises in selection of patients. One included health care center in a multicultural area unfortunately had to be excluded from the trial due to language problems and following methodological errors. Also, an inclusion criterion was to understand Swedish. Consequently, we did not achieve a diversity in terms of ethnicity and cultural background. The participants were comparable with the Swedish hypertension population in terms of age, but the sex distribution differed, with a majority of men in our study [41]. As always in studies like this, there is a risk for selection bias in the recruitment of patients. The patients who are already aware of their health status and motivated to treat their condition may have chosen to participate in the trial to a higher degree.

Conclusion

Using technology for strengthening patients' potential for self-management has the possibility to change the relationship between patients and professionals. The patients perceived themselves as more active and motivated in their BP treatment. When using the system as intended, the professionals experienced it as a resource for communication regarding BP and lifestyle. Both patients and professionals described a consultation on more equal grounds, laying the foundation for a constructive partnership.

To realize the potential in a system like this, health care professionals need to be motivated and interested in new approaches in management of chronic conditions. Integration of the technology in the existing technical system is essential. Health care professionals also need to receive a thorough introduction so that they, in turn, can properly instruct and motivate patients to use the system and read the manual. If this is not achievable, introduction of a new technical solution may instead increase workload and become a burden in chronic condition management both for professionals and patients.

Acknowledgments

The authors sincerely thank all the patients and professionals who participated in this qualitative study. This work was supported by grant 20170102 from the Kamprad Foundation, grant 20170251 from the Heart and Lung Foundation, and grant 2018-02648 from the Swedish Research Council.

Authors' Contributions

All authors contributed to the design and planning of the study. UA and UB performed the interviews. UA, UB, and KK contributed to the analysis and interpretation of data. UA drafted the manuscript. UA, KK, AR, and PM critically revised and gave final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codes for focus group interview with patients.
[PDF File (Adobe PDF File), 175 KB - jmir_v23i6e26143_app1.pdf]

References

1. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, ESC Scientific Document Group. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J 2018 Sep 01;39(33):3021-3104. [doi: 10.1093/eurheartj/ehy339] [Medline: 30165516]
2. Piepoli M, Hoes A, Agewall S, Albus C, Brotons C, Catapano A, ESC Scientific Document Group. 2016 European guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2016 Aug 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
3. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
4. Holman H, Lorig K. Patients as partners in managing chronic disease: partnership is a prerequisite for effective and efficient health care. BMJ 2000 Feb 26;320(7234):526-527 [FREE Full text] [doi: 10.1136/bmj.320.7234.526] [Medline: 10688539]
5. Thulesius H. Increased importance of digital medicine and eHealth during the Covid-19 pandemic. Scand J Prim Health Care 2020 Jun 02;38(2):105-106. [doi: 10.1080/02813432.2020.1770466] [Medline: 32484725]
6. Shahaj O, Denneny D, Schwappach A, Pearce G, Epiphaniou E, Parke HL, et al. Supporting self-management for people with hypertension: a meta-review of quantitative and qualitative systematic reviews. J Hypertens 2019 Feb;37(2):264-279. [doi: 10.1097/HJH.0000000000001867] [Medline: 30020240]

7. CEN/TC 450: patient involvement in person-centred care. European Committee for Standardization. 2020. URL: https://standards.cen.eu/dyn/www/f?p=204:7:0::::FSP_ORG_ID:2151116&cs=1D44A19AFA1BBECA019313BDB1377FF73 [accessed 2021-05-19]
8. Ekman I, Swedberg K, Taft C, Lindseth A, Norberg A, Brink E, et al. Person-centered care—ready for prime time. Eur J Cardiovasc Nurs 2011 Dec;10(4):248-251. [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
9. Hook ML. Partnering with patients—a concept ready for action. J Adv Nurs 2006 Oct;56(2):133-143. [doi: 10.1111/j.1365-2648.2006.03993.x] [Medline: 17018062]
10. Gallant MH, Beaulieu MC, Carnevale FA. Partnership: an analysis of the concept within the nurse-client relationship. J Adv Nurs 2002 Oct;40(2):149-157. [doi: 10.1046/j.1365-2648.2002.02357.x] [Medline: 12366644]
11. Yuan S, Murphy J. Partnership in nursing care: a concept analysis. TMR Integr Nurs 2019;3(1):21-26. [doi: 10.12032/TMRIN20181029]
12. Wolf A, Moore L, Lydahl D, Naldemirci Ö, Elam M, Britten N. The realities of partnership in person-centred care: a qualitative interview study with patients and professionals. BMJ Open 2017 Jul 17;7(7):e016491 [FREE Full text] [doi: 10.1136/bmjopen-2017-016491] [Medline: 28716793]
13. Wildevuur SE, Simonse LW, Groenewegen P, Klink A. Information and communication technology enabling partnership in person-centred diabetes management: building a theoretical framework from an inductive case study in The Netherlands. BMJ Open 2019 Jun 16;9(6):e025930 [FREE Full text] [doi: 10.1136/bmjopen-2018-025930] [Medline: 31209085]
14. Hallberg I, Taft C, Ranerup A, Bengtsson U, Hoffmann M, Höfer S, et al. Phases in development of an interactive mobile phone-based system to support self-management of hypertension. Integr Blood Press Control 2014 May;7:19-28 [FREE Full text] [doi: 10.2147/IBPC.S59030] [Medline: 24910510]
15. Bengtsson U, Kasperowski D, Ring L, Kjellgren K. Developing an interactive mobile phone self-report system for self-management of hypertension. Part 1: patient and professional perspectives. Blood Press 2014 Oct;23(5):288-295 [FREE Full text] [doi: 10.3109/08037051.2014.883203] [Medline: 24564289]
16. Bengtsson U, Kjellgren K, Höfer S, Taft C, Ring L. Developing an interactive mobile phone self-report system for self-management of hypertension. Part 2: content validity and usability. Blood Press 2014 Oct;23(5):296-306 [FREE Full text] [doi: 10.3109/08037051.2014.901009] [Medline: 24786778]
17. Ranerup A, Hallberg I. Actors and intentions in the development process of a mobile phone platform for self-management of hypertension. Inform Health Soc Care 2014 Jun 24:1-20. [doi: 10.3109/17538157.2014.924948] [Medline: 24960437]
18. Lundin M, Mäkitalo Å. Co-designing technologies in the context of hypertension care: negotiating participation and technology use in design meetings. Inform Health Soc Care 2017 Jan 18;42(1):18-31. [doi: 10.3109/17538157.2015.1113176] [Medline: 26890491]
19. Hallberg I, Ranerup A, Kjellgren K. Supporting the self-management of hypertension: patients' experiences of using a mobile phone-based system. J Hum Hypertens 2016 Feb;30(2):141-146 [FREE Full text] [doi: 10.1038/jhh.2015.37] [Medline: 25903164]
20. Bengtsson U, Kjellgren K, Hallberg I, Lindwall M, Taft C. Improved blood pressure control using an interactive mobile phone support system. J Clin Hypertens (Greenwich) 2016 Feb;18(2):101-108 [FREE Full text] [doi: 10.1111/jch.12682] [Medline: 26456490]
21. Hallberg I, Ranerup A, Bengtsson U, Kjellgren K. Experiences, expectations and challenges of an interactive mobile phone-based system to support self-management of hypertension: patients' and professionals' perspectives. PPA 2018 Mar;Volume 12:467-476. [doi: 10.2147/ppa.s157658]
22. Bengtsson U, Kjellgren K, Hallberg I, Lundin M, Mäkitalo Å. Patient contributions during primary care consultations for hypertension after self-reporting via a mobile phone self-management support system. Scand J Prim Health Care 2018 Mar 18;36(1):70-79 [FREE Full text] [doi: 10.1080/02813432.2018.1426144] [Medline: 29343156]
23. Midlöv P, Nilsson PM, Bengtsson U, Hoffmann M, Wennersten A, Andersson U, et al. PERson-centredness in hypertension management using information technology (PERHIT): a protocol for a randomised controlled trial in primary health care. Blood Press 2020 Jun 09;29(3):149-156. [doi: 10.1080/08037051.2019.1697177] [Medline: 31814476]
24. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
25. Thangada ND, Garg N, Pandey A, Kumar N. The emerging role of mobile-health applications in the management of hypertension. Curr Cardiol Rep 2018 Jul 26;20(9):78. [doi: 10.1007/s11886-018-1022-7] [Medline: 30046971]
26. Omboni S, Caserini M, Coronetti C. Telemedicine and mHealth in hypertension management: technologies, applications and clinical evidence. High Blood Press Cardiovasc Prev 2016 Sep;23(3):187-196. [doi: 10.1007/s40292-016-0143-6] [Medline: 27072129]
27. Palmer MJ, Barnard S, Perel P, Free C. Mobile phone-based interventions for improving adherence to medication prescribed for the primary prevention of cardiovascular disease in adults. Cochrane Database Syst Rev 2018 Dec 22;6:CD012675. [doi: 10.1002/14651858.CD012675.pub2] [Medline: 29932455]

28. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile apps to support the self-management of hypertension: systematic review of effectiveness, usability, and user satisfaction. JMIR Mhealth Uhealth 2018 Jul 23;6(7):e10723 [FREE Full text] [doi: 10.2196/10723] [Medline: 30037787]
29. Mohammadi R, Ayatolahi M, Hoveidamanesh S, Ghanavati R, Pournik O. Reflection on mobile applications for blood pressure management: a systematic review on potential effects and initiatives. Stud Health Technol Inform 2018;247:306-310. [Medline: 29677972]
30. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. J Am Med Inform Assoc 2020 Jun 01;27(6):939-945 [FREE Full text] [doi: 10.1093/jamia/ocaa029] [Medline: 32357368]
31. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
32. Krueger RA. Focus Groups: A Practical Guide for Applied Research. 5th Edition. Thousand Oaks: Sage Publications; 2014.
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Castleberry A, Nolen A. Thematic analysis of qualitative research data: is it as easy as it sounds? Curr Pharma Teach Learn 2018 Jun;10(6):807-815. [doi: 10.1016/j.cptl.2018.03.019] [Medline: 30025784]
35. Wildevuur S, Thomese F, Ferguson J, Klink A. Information and communication technologies to support chronic disease self-management: preconditions for enhancing the partnership in person-centered care. J Participat Med 2017 Jul 07;9(1):e14. [doi: 10.2196/jopm.8846]
36. Kim JY, Wineinger NE, Steinhubl SR. The influence of wireless self-monitoring program on the relationship between patient activation and health behaviors, medication adherence, and blood pressure levels in hypertensive patients: a substudy of a randomized controlled trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
37. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (sms-text adherence support [STAR]): a single-blind, randomized trial. Circulation 2016 Feb 09;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]
38. Bolmsjö B, Wolff M, Nymberg VM, Sandberg M, Midlöv P, Calling S. Text message-based lifestyle intervention in primary care patients with hypertension: a randomized controlled pilot trial. Scand J Prim Health Care 2020 Sep 20;38(3):300-307 [FREE Full text] [doi: 10.1080/02813432.2020.1794392] [Medline: 32686972]
39. Kjellgren KI, Svensson S, Ahlner J, Säljö R. Antihypertensive treatment and patient autonomy: the follow-up appointment as a resource for care. Pat Educ Couns 2000 Apr;40(1):39-49. [doi: 10.1016/s0738-3991(99)00042-7]
40. Samkange-Zeeb F, Borisova L, Padilla B, Bradby H, Phillimore J, Zeeb H, et al. Superdiversity, migration and use of internet-based health information: results of a cross-sectional survey conducted in 4 European countries. BMC Public Health 2020 Aug 20;20(1):1263 [FREE Full text] [doi: 10.1186/s12889-020-09329-6] [Medline: 32819317]
41. Hasselström J, Zarrinkoub R, Holmquist C, Hjerpe P, Ljungman C, Qvarnström M, et al. The Swedish Primary Care Cardiovascular Database (SPCCD): 74,751 hypertensive primary care patients. Blood Press 2014 Apr;23(2):116-125. [doi: 10.3109/08037051.2013.814829] [Medline: 23914944]

Abbreviations

BP: blood pressure
COREQ: Consolidated Criteria for Reporting Qualitative Research
mHealth: mobile health
PCC: person-centered care
PERHIT: Person-Centeredness in Hypertension Management Using Information Technology
PHCC: primary health care center
RCT: randomized controlled trial

Edited by R Kukafka; submitted 30.11.20; peer-reviewed by T Alessa, O Blanson Henkemans; comments to author 19.01.21; revised version received 15.02.21; accepted 11.04.21; published 03.06.21.

Please cite as:
Andersson U, Bengtsson U, Ranerup A, Midlöv P, Kjellgren K
Patients and Professionals as Partners in Hypertension Care: Qualitative Substudy of a Randomized Controlled Trial Using an Interactive Web-Based System Via Mobile Phone
J Med Internet Res 2021;23(6):e26143
URL: https://www.jmir.org/2021/6/e26143
doi:10.2196/26143
PMID:34081021

Original Paper

Fidelity of Delivery and Contextual Factors Influencing Children's Level of Engagement: Process Evaluation of the Online Remote Behavioral Intervention for Tics Trial

Kareem Khan[1], BSc, MSc; Chris Hollis[1,2,3], BSc, MBBS, DCH, PhD, FRCPsych; Charlotte L Hall[1], BSc, MSc, PhD; Elizabeth Murray[4], BA, MSc, PhD; E Bethan Davies[1,2], BSc, MSc, PhD; Per Andrén[5], BSc, MSc; David Mataix-Cols[5], BSc, MSc, PhD; Tara Murphy[6], BSc, MSc, DClinPsy, QiCN; Cris Glazebrook[1,2], RGN, PhD, CPsychol

[1]Division of Psychiatry and Applied Psychology, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[2]NIHR MindTech Medtech Co-operative, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[3]NIHR Nottingham Biomedical Research Centre, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[4]Research Department of Primary Care and Population Health, University College London, London, United Kingdom

[5]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden

[6]Tic Disorder Clinic, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Kareem Khan, BSc, MSc
Division of Psychiatry and Applied Psychology
School of Medicine
University of Nottingham
Institute of Mental Health, Jubilee Campus
University of Nottingham Innovation Park, Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 0115 823 1294
Email: kareem.khan@nottingham.ac.uk

Abstract

Background: The Online Remote Behavioral Intervention for Tics (ORBIT) study was a multicenter randomized controlled trial of a complex intervention that consisted of a web-based behavioral intervention for children and young people with tic disorders. In the first part of a two-stage process evaluation, we conducted a mixed methods study exploring the reach, dose, and fidelity of the intervention and contextual factors influencing engagement.

Objective: This study aims to explore the fidelity of delivery and contextual factors underpinning the ORBIT trial.

Methods: Baseline study data and intervention usage metrics from participants in the intervention arm were used as quantitative implementation data (N=112). The experiences of being in the intervention were explored through semistructured interviews with children (n=20) and parent participants (n=20), therapists (n=4), and referring clinicians (n=6). A principal component analysis was used to create a comprehensive, composite measure of children and young people's engagement with the intervention. Engagement factor scores reflected relative uptake as assessed by a range of usage indices, including chapters accessed, number of pages visited, and number of log-ins. The engagement factor score was used as the dependent variable in a multiple linear regression analysis with various contextual variables as independent variables to assess if there were any significant predictors of engagement.

Results: The intervention was implemented with high fidelity, and participants deemed the intervention acceptable and satisfactory. The engagement was high, with child participants completing an average of 7.5 of 10 (SD 2.7) chapters, and 88.4% (99/112) of participants completed the minimum of the first four chapters—the predefined threshold effective dose. Compared with the total population of children with tic disorders, participants in the sample tended to have more educated parents and lived in more economically advantaged areas; however, socioeconomic factors were not related to engagement factor scores. Factors associated with higher engagement factor scores included participants enrolled at the London site versus the Nottingham site (*P*=.01), self-referred versus clinic referred (*P*=.04), higher parental engagement as evidenced by the number of parental chapters completed (n=111; ρ=0.73; $P<.001$), and more therapist time for parents (n=111; ρ=0.46; $P<.001$). A multiple linear regression indicated that parents' chapter completion (β=.69; $t_{110}=10.18$; $P<.001$) and therapist time for parents (β=.19; $t_{110}=2.95$; $P=.004$) were the only significant independent predictors of child engagement factor scores.

Conclusions: Overall, the intervention had high fidelity of delivery and was evaluated positively by participants, although reach may have been constrained by the nature of the randomized controlled trial. Parental engagement and therapist time for parents were strong predictors of intervention implementation, which has important implications for designing and implementing digital therapeutic interventions in child and adolescent mental health services.

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-019-3974-3

(J Med Internet Res 2021;23(6):e25470) doi:10.2196/25470

KEYWORDS

process evaluation; implementation fidelity; Tourette syndrome; chronic tic disorders; online behavioral intervention; mixed methods; children and young people

Introduction

Tics are sudden, brief, rapid, and recurrent nonrhythmic movements or vocalizations that are more common in children and young people (CYP) than in adults [1] and more prevalent in males than in females [2]. Tic onset typically occurs between the ages of 3 and 8 years (mean onset is 6 and 7 years of age) [3], with the reported average age of greatest tic severity by the age of 10 years [4]. Although most CYP with tics only require educational support as the main form of treatment [5], there are interventions available for severe or disabling tics, as in Tourette syndrome or chronic tic disorders. Historically, pharmacotherapy, such as antipsychotics, has been the first line of treatment for severe tics; however, they often have undesirable side effects, such as weight gain and sleepiness [6].

Behavioral interventions are appealing and effective alternatives to pharmacotherapy. However, they require the patient to invest time and energy in practicing demanding behavioral techniques, such as tic control or habit reversal. Despite the benefits and evidence-based effectiveness of behavioral therapies for tic disorders [7-9], there is great difficulty in patients accessing behavioral treatments because of a shortage of trained therapists [10]. A promising development in increasing accessibility to behavioral treatments is the use of digital health interventions (DHIs) [11]. Preliminary evidence suggests that DHIs are efficacious for CYP with tic disorders in pilot randomized controlled trials (RCTs) [12-14]. A study that has assessed DHIs for tic disorders is the *Online Remote Behavioral Intervention for Tics* (ORBIT) trial, which has been described in detail previously [15] (see Textbox 1 for a brief description).

Textbox 1. Brief description of the ORBIT (Online Remote Behavioral Intervention for Tics) trial.

Design: A 10-week, 2-armed, parallel-group, single-blind randomized controlled trial with an embedded process evaluation.

Aim: To evaluate the effectiveness of a web-based, remote, therapist-supported, and parent-guided behavioral intervention for tics, initially developed and piloted in Sweden.

Intervention group: A total of 112 children and young people received 10 modules (called "chapters") of behavioral therapy following the principles of exposure and response prevention via a secure web-based platform, with access to a therapist, delivered over a period of 10-12 weeks.

Control group: A total of 112 children and young people received 10 chapters of psychoeducation via a secure web-based platform, with access to a therapist, delivered over a period of 10-12 weeks.

Primary outcome: Total Tic Severity Score on the Yale Global Tic Severity Scale at 3 months postrandomization.

Therapist role: Both children and parents had regular contact with a therapist during the 10-12 week period via messages that were sent within the treatment platform (resembling an email) or via telephone, if required. The therapist was also able to directly comment on exercises that the participant had been working on and give specific feedback to motivate participants.

Parent role: One or both of the child's parents received a separate log-in to the web-based treatment platform, where they could access their own chapters. The parent chapters contained information regarding parent coping strategies and how to support their child. The parents also had access to the assigned therapist.

The population impact of any given intervention depends on both its effectiveness and its reach, defined as the proportion of the target population who access the intervention [16]. Although RCTs are the *gold standard* method for determining efficacy, additional data are needed before deciding whether an intervention should be adopted into mainstream health care. These additional data include understanding the reach of the intervention and the extent to which the data from an RCT, where the delivery of the intervention is often tightly controlled and monitored, can be extrapolated to use in routine health care.

It has been argued that studies addressing questions about reach and effectiveness in routine care are needed [17,18]. However, such studies are expensive, and a process evaluation conducted alongside an RCT is an efficient method of maximizing the information yielded by the trial.

The Medical Research Council (MRC) has developed specific guidelines for conducting process evaluations of complex interventions [19]. A complex intervention is defined as an intervention with several interacting components [20], and the MRC outlines three essential components for evaluating complex interventions: implementation, mechanisms of impact, and context. Implementation can refer to how an intervention will be delivered within routine clinical practice, having shown efficacy in an outcome evaluation. However, this paper is concerned with another aspect of implementation: the extent to which the delivery of an intervention is achieved within the context of an RCT and the structures and processes through which an intervention is delivered as intended (ie, fidelity) [19]. For complex interventions such as DHIs, an important component of implementation fidelity is the degree to which participants engage with the intervention and use it as intended. Effective engagement requires participants to register with the program and then continue to use it and apply the recommended behavioral techniques over time. The nonuse of DHIs is a well-recognized challenge (eg, the Eysenbach Law of Attrition [21]) and can be considered in two parts: initial uptake (eg, registration and onboarding) and ongoing engagement.

To evaluate intervention implementation, the MRC guidelines for process evaluations suggest researchers assess (1) reach—the extent to which a target audience comes into contact with the intervention; (2) dose—how much of the intervention was delivered and received; (3) fidelity—the quality of what was delivered; and (4) adaptations—any modifications made to an intervention to achieve better contextual fit. The intended target audience for ORBIT was CYP with tic disorders; however, pertinent questions could be asked, such as whether there were socioeconomic biases in who was reached. In terms of dosage, the ORBIT protocol [15] states that the intervention should consist of 10 individual intervention chapters following a suggested frequency and a total duration of 10-12 weeks. The first four chapters delivered core content, including learning about tics and practicing suppressing tics. Completing these four chapters was designated as the minimum *dose* required for treatment completion by the ORBIT clinical team. There were six additional chapters offering reinforcement, further practice, and relapse prevention. For DHIs, the fidelity of delivery of the intervention is ensured by the web-based delivery platform. However, the intervention experienced by the user is highly dependent on the extent to which they engage with the intervention and use it as intended and the quality of delivery [22]. Hence, in this paper, we look at usage and the proportion of participants receiving the predefined *minimum effective dose* of four or more chapters. Finally, understanding adaptations to the intended intervention involves exploring whether they improve the contextual fit or compromise the functioning of the intervention [23] or whether they represent innovation or intervention drift [24]. Participants were able to modify various intervention components, such as the *tic stopwatch*, which was used to self-time the length of tic control. This study aims to conduct the first part of a two-stage process evaluation of the ORBIT trial outlined in the study protocol [25]. Part 1 focuses on intervention implementation by exploring the fidelity of delivery experienced by participants using usage statistics, reach, and acceptability of the intervention. It also investigates contextual factors associated with the observed variation in uptake and usage by examining the components specified in the MRC guidelines [19]. Part 2 will focus on the impact mechanisms, and the findings will be reported in a future article. Table 1 shows the two parts of the process evaluation: areas of research, explanatory data, and outcomes.

Table 1. Process evaluation parts, areas of research, explanatory data, and outcomes.

Research questions	Process evaluation components	Explanatory data	Outcomes
Part 1. Intervention implementation (What is implemented and how?)	• Fidelity of implementation • Dose of intervention delivered • Adaptations • Reach	• Therapist contact (N=112) • Intervention engagement (N=112) • Usage metrics (N=112) • Clinician (n=6), children and parent (n=20), and therapist (n=4) interviews	• Engagement and satisfaction with the intervention
Part 2. Mechanisms of impact (How does it produce change?)	• Mediators and moderators • Unexpected pathways and consequences	• Usage metrics • Therapist contacts • Clinician, children and parent, and therapist interviews	• Relationship between engagement with the intervention and change in tic severity
Part 1 and 2. Context (How do factors external to the intervention affect intervention implementation and change?)	• Factors related to fidelity of delivery (part 1) and improvement in tics (part 2)	• Demographic data • Clinician, children and parent, and therapist interviews • Comorbidities • Baseline severity of tics	• Engagement with the intervention (part 1) • Change in tic severity (part 2)

Methods

Study Design

This study followed the MRC guidelines [19] for the process evaluation of complex interventions. It used a mixed methods, longitudinal design to explore the implementation fidelity of a web-based intervention for CYP with tics [15] and the contextual factors that influenced the level of engagement.

Participants

The sample included in the quantitative phase of the process evaluation consisted of key information from all participants (N=112) from the intervention arm of the RCT. The sample included in the qualitative component of the process evaluation consisted of interviews with children and parent participants (target n≥20), interviews with all therapists delivering the intervention or supervising the therapists, and interviews with referring clinicians (target n>5).

Quantitative Data Collection

Quantitative process data were collected simultaneously with enrollment, intervention delivery, and outcome data collection in the main RCT.

Demographic and Clinical Data

Demographic and clinical information were recorded using a baseline demographic questionnaire. These data included the child's age, residence (full postcode), gender, ethnicity, parental education level and occupation, all current suspected or confirmed diagnoses and interventions, and medication use.

Index of Multiple Deprivation

The index of multiple deprivation (IMD; 2019) is a relative measure of deprivation across seven different domains: income deprivation; employment deprivation; education, skills, and training deprivation; health deprivation and disability; crime; barriers to housing and services; and living environment deprivation [26]. On the basis of the 6-digit zip code, the rank of deprivation associated with participants' area of residence was calculated from 32,844 small areas or neighborhoods in England, with higher ranks indicating greater deprivation. Ranks were recoded into quintiles, with 1 being the most deprived and 5 being the least deprived.

Yale Global Tic Severity Scale

The primary outcome measure used in the ORBIT intervention was the Total Tic Severity Score (TTSS), as measured by the Yale Global Tic Severity Scale (YGTSS). The YGTSS is a valid and reliable clinician-rated scale [27], which scores the severity of motor and vocal tics separately by evaluating the number, frequency, intensity, complexity, and interference of tics. Each domain was scored on a 0-5 scale, and 2 tic severity scores were given: total motor (0-25) score and total vocal (0-25) score, which when combined give the TTSS (0-50).

Mood and Feelings Questionnaire

The Mood and Feelings Questionnaire (MFQ) [28] is a 33 ‐ item measure evaluating depressive symptoms rated on a 3 ‐ point scale: 0 is *not true*, 1 is *sometimes*, and 2 is *true*. Total scores range from 0 to 66, with higher scores reflecting more severe depression. A cutoff score of ≥29 is generally used to suggest clinically significant depression [29].

Usage Metrics

Web usage data were collected and recorded from the participants throughout the trial. This included the following measures: number of chapters completed per child and per parent, total therapists' time per child and per parent, individual therapist's telephone time with participants, the volume of written communication (total number of characters) submitted by child and parent via the web-based system, total number of log-ins for child and parent, average time between each log-in (in days) for child and parent, and average pages visited per log-in for child and parent.

Satisfaction and Treatment Credibility

At the 3-week postrandomization point of treatment, all participants were asked to rate treatment credibility; two questions were asked: one relating to how well suited the participant felt the intervention was for helping CYP to manage their tics and the other question was about how much better they expected to feel as a result of the intervention. The responses were rated on a Likert scale of 0 to 4 for each question, with higher scores indicating higher treatment credibility. At the primary endpoint, all participants were asked to rate their satisfaction with the intervention. In total, eight satisfaction questions were asked with responses rated on a scale of 0-4, meaning that the overall satisfaction score was out of 32.

Qualitative Data Collection

Interviews with therapists and therapist supervisors involved in the ORBIT trial were conducted early in the study and near the end of recruitment to gain an understanding of their experience at different time points. Interviews with referring clinicians were conducted at the end of recruitment. Interviews with children participants and one of their parents were conducted following completion of the intervention at the 3-month (primary endpoint) follow-up assessment in the main RCT to minimize the risk of bias in the outcomes. Recruitment for the interviews began in August 2018 and ended in October 2019.

All interviews were conducted face-to-face, via telephone, or via videoconference (WebEx or Skype). Younger children were interviewed together with their parents, whereas older children (eg, aged >13 years) were interviewed separately. Participants were purposively sampled so that a diverse range of views on the intervention were voiced [30]. This included ensuring that perspectives were heard from participants with a range of ages, gender, ethnicity, and level of interaction with the intervention. The overall sample enabled diverse intervention views and ensured that the data reached the saturation level [31]. In addition to the interviews, at the end of the treatment, all participants were asked to give their overall feedback on the intervention, to which they could provide open-ended responses. Table 2 demonstrates how various data sources contributed to the different components of implementation fidelity.

Table 2. Implementation fidelity components and data sources.

Data sources	Reach	Dose	Fidelity	Adaptations	Context
Quantitative data sources					
Demographic and clinical data	✓[a]				✓
Usage metrics		✓			
Treatment credibility and satisfaction			✓		
Qualitative data sources					
Child interviews		✓	✓	✓	✓
Parent interviews		✓	✓		✓
Therapist and clinician interviews	✓		✓	✓	✓
End-of-treatment feedback questionnaire			✓		

[a] Associated data with implementation component.

Data Analysis

Data were tested for normality using the Kolmogorov–Smirnov test. A principal component analysis was run to determine a composite measure of the level of engagement. Correlations between variables were examined using bivariate Spearman correlations, and a *t* test was performed to explore any significant differences between groups using chi-square tests to explore the differences between categorical variables. Multiple linear regression was used to identify predictors of engagement with the independent variables. All statistical analyses used a significance level of $P<.05$ and were conducted using SPSS Statistics 27 (IBM Corporation).

All interviews were recorded by using either videoconferencing software or a Dictaphone and were then transcribed verbatim. Transcripts were checked for accuracy against the recordings, with any corrections made as appropriate and anonymized for confidentiality purposes. As the process evaluation was a combination of exploration and description, the framework method [32] of analysis was used to identify, analyze, and report patterns within the transcribed interviews. Moreover, the steps outlined by Gale et al [33] were systematically followed to create an overall framework matrix using categories of engagement and potential moderators. Consistency of analysis was ensured through the use of a codebook and frequent meetings between researchers. Researcher bias was minimized through regular cross-checking of data and outcomes by the members of the research team.

The NVivo 12 software package (QSR International) was used to analyze the interview data. In addition, the end-of-treatment feedback questionnaire was exported to a Microsoft Excel spreadsheet, and quantitative content analysis [34] was performed. Overall, the findings from the qualitative analysis were linked to relevant quantitative engagement outcomes and contextual factors to assess which potential moderators may have influenced implementation fidelity and in what way, in an approach termed *triangulation* [35].

Availability of Data and Materials

Some of the data generated or analyzed during this study are included in this paper and its supplementary information files or are available from the corresponding author upon reasonable request. The full data sets generated or analyzed during this study are also available from the corresponding author upon reasonable request.

Ethical Considerations

Ethical approval for the study was obtained from the North West Greater Manchester Central Research Ethics Committee (REC: 18/NW/0079). We sought written parental consent and written informed assent or consent to participate in the study from CYP. This covered the process evaluation measures. All participants provided verbal consent to be audio-recorded for all interviews.

Results

Overview of Qualitative Sample

Semistructured interviews were conducted with children (n=20), parents (n=20), therapists (n=4), and clinicians (n=6). The average age of the child interviewees was 12 years (SD 2.1); range 9-16 years), of which 80% (16/20) were male and 20% (4/20) were female. Most participants were White (18/20, 90%). The mean TTSS was 28.8 out of 50 (SD 7.2), with a range of 13-45 for the child interviewees. All 20 interviews with the parents were with the CYP's mothers, with all 20 having completed at least some further education. One of the therapist interviewees was a therapist's supervisor, and half of the clinicians (3/6, 50%) were consultant psychiatrists.

Reach

Participants were eligible for the study if they were aged between 9 and 17 years and competent to provide written, informed consent (parental consent for a child aged <16 years), had a suspected or confirmed tic disorder (as confirmed by scores on the YGTSS), and had broadband internet access and regular use of a computer, with mobile phone text messaging facilities. Patients were excluded from the study if they had received any form of structured behavioral intervention for tics within the preceding 12 months, had a change of medication for tics within the previous 2 months, had any diagnoses of alcohol or substance dependence, psychosis, suicidality, anorexia nervosa, or moderate or severe intellectual disability, were an immediate risk to self or others, and/or parents or children were not able to speak, read, or write English.

A total of 445 families expressed an interest in participating in the study either through self-referral via the *Tourettes Action* charity website (n=251) or clinic referral (n=194); however, 47 were subsequently uncontactable, and 90 were ineligible to participate. Of the 308 potentially eligible CYP, 84 (27.3%) families declined to participate, and 50% (112/224) of CYP (90 males and 22 females) with a mean age of 12.2 (SD 2) years were randomized to the intervention arm of the ORBIT trial and included in the process evaluation. The sample was predominantly White (96/112, 85.7%) and well-educated, with more than half (60/112, 53.5%) of the participants' mothers having completed university or higher education.

The median IMD rank was 19,318 and ranged from 147 to 32,668 (out of 32,844). Of the 112 participants, 8 (7.1%) were in the most deprived quintile 1, 31 (27.7%) in quintile 2, 18 (16.1%) in quintile 3, 26 (23.3%) in quintile 4, and 29 (25.6%) in the least deprived quintile 5. Although the reach of the intervention was not limited geographically, for research purposes, participants had to attend a baseline screening assessment at either the Nottingham study site (57/112, 50.9%) or the London study site (55/112, 49.1%) depending on personal preference and/or location of residence. All participants were based in England, with 56.3% (63/112) of participants living in towns, 26.7% (30/112) living in cities, and 16.9% (19/112) living in villages.

In terms of clinical characteristics, the intervention reached a moderately severe symptomatic sample with a mean TTSS of 28.4 out of 50 (SD 7.7) ranging from 12 to 50. Most participants (98/112, 87.5%) were not on any medication for their tics, and less than half of the overall intervention sample had no diagnosed or suspected comorbidities (51/112, 45.5%). Among those who had a comorbid diagnosis, the most common was anxiety disorder (34/112, 30.4%), followed by attention-deficit/hyperactivity disorder (26/112, 23.2%). An assessment of depressive symptoms by the MFQ showed a mean score of 16.3 out of 66 (SD 11.3), with 12.5% (14/112) of participants scoring above the cut-off (≥29), suggesting clinically significant depression [29] (Table 3).

It was not possible to interview people who had not taken part in the study, so the qualitative data threw little light on reach. However, a clinician identified that some families were worried about the level of commitment involved and associated travel to one of the study sites (quote 1) under the theme *clinician perceptions of and contribution to recruitment* (see Multimedia Appendix 1 for a full list of framework categories and themes and Multimedia Appendix 2 for a full list of quotes). Another clinician highlighted the lack of access to children with intellectual disabilities (quote 2). Finally, one of the clinicians struggled to gain her colleagues' interest in the intervention despite numerous attempts (quote 3).

Table 3. Demographic and clinical characteristics of participants in the Online Remote Behavioral Intervention for Tics trial intervention group (N=112).

Variable	Intervention group
Gender, n (%)	
Male	90 (80.4)
Female	22 (19.6)
Study site, n (%)	
Nottingham	57 (50.9)
London	55 (49.1)
Ethnicity, n (%)	
White	96 (85.7)
Asian	7 (6.2)
Mixed race	3 (2.7)
Other	6 (5.4)
Supporter, n (%)	
Mother	93 (83)
Father	16 (14.3)
Other	3 (2.7)
Highest level of education (mother), n (%)	
Did not complete compulsory education	3 (2.7)
Completed compulsory secondary education	16 (14.3)
Completed further education	33 (29.5)
Completed university or higher education	43 (38.4)
Completed postgraduate taught degree	11 (9.7)
Completed doctorate or medical degree	6 (5.4)
Highest level of education (father), n (%)	
Did not complete compulsory education	2 (1.8)
Completed compulsory secondary education	29 (25.9)
Completed further education	35 (31.2)
Completed university or higher education	29 (25.9)
Completed postgraduate taught degree	10 (8.9)
Completed doctorate or medical degree	7 (6.3)
Method of referral, n (%)	
Self	69 (61.6)
Clinic	43 (38.4)
Age (years), mean (SD)	12.2 (2)
IMD[a] rank, median (range)	19318 (147-32,668)
No tic medication, n (%)	98 (87.5)
On tic medication, n (%)	14 (12.5)
Comorbidities, n (%)	61 (54.5)
No comorbidities, n (%)	51 (45.5)
TTSS[b] baseline score, mean (SD)	28.4 (7.7)
MFQ[c], mean (SD)	16.3 (11.3)

[a] IMD: index of multiple deprivation.

[b] TTSS: Total Tic Severity Score.

[c] MFQ: Mood and Feelings Questionnaire.

Dose

Child participants completed an average of 7.5 (SD 2.7; Table 4) chapters, and their parents completed an average of 7.6 (SD 2.8; Table 5) out of 10 chapters of the intervention, indicating high engagement. Only 11.6% (13/112) of child participants and 15.2% (17/112) of parents failed to meet the criteria for treatment completion (ie, the minimum of the first four chapters completed as per protocol) with a total of 88.4% (99/112) of child participants and 84.8% (95/112) of parents completing their treatment, meaning that adherence to the intervention was high. Indeed, 41% (46/112) of CYP and 46.4% (52/112) of parents completed all 10 intervention chapters, and only 1 child participant failed to complete any chapters. Participants were given 10 weeks of supported therapeutic input to complete their treatment chapters. In some circumstances, such as holidays or particularly busy periods, 1 or 2 weeks were added to supplement this time. Although most families (73/112, 65.2%) completed their therapy within 10 weeks, 34.8% (39/112) required extra time to complete treatment. Child participants logged onto the web-based treatment platform an average of 19.8 (SD 10.9) times throughout the 10-12 weeks, with an average of 4.2 (SD 2.6) days between log-ins. In terms of total interactions with their assigned therapist, child participants required their therapist's assistance on the web for an average of 59 minutes and 14 seconds (SD 29 min and 08 s) throughout treatment, resulting in approximately 6 minutes per child each week. In contrast, parents interacted on the web with their assigned therapist for an average of 1 hour, 23 minutes, and 55 seconds (SD 42 min and 45 s), resulting in approximately 8 minutes per parent each week. Of the 112 CYP, only 2 (1.8%) were contacted by telephone by their assigned therapist, whereas of the 112 parents, 49 (43.7%) were contacted by telephone by their assigned therapist.

Interview data relating to participants' *perceptions of the ORBIT organization* covered the implementation component of the dose. Although most participants felt that the intervention was just the right length, some CYP wished to have a longer time to access their therapist (quote 4). A child felt that the intervention could have been condensed to make it shorter (quote 5). On the whole, parents agreed with their child that the dose received was just right, with a parent claiming that if it were longer, it would have negatively affected engagement levels (quote 6).

Table 4. Usage data for child participants in the Online Remote Behavioral Intervention for Tics trial intervention group (N=112).

Variable	Value	
	Median (range)	Mean (SD)
Chapters completed	8 (0-10)	7.5 (2.7)
Total therapist time (hh:mm:ss)	00:53:57 (00:07:27-03:11:08)	00:59:14 (00:29:08)
Telephone time with therapist (hh:mm:ss)	00:00:00 (00:00:00-00:18:44)	00:00:10 (00:01:46)
Number of log-ins	19 (3-57)	19.8 (10.9)
Number of days between log-ins	3 (1-16)	4.2 (2.6)
Number of pages visited per log-in	15 (7-38)	16.9 (5.8)
Total number of characters submitted	2507 (238-8749)	2784 (1608)

Table 5. Usage data for parents in the Online Remote Behavioral Intervention for Tics trial intervention group (N=112).

Variable	Value	
	Median (range)	Mean (SD)
Chapters completed	9 (1-10)	7.6 (2.8)
Total therapist time (hh:mm:ss)	01:15:33 (00:22:01-04:48:19)	01:23:55 (00:42:45)
Telephone time with therapist (hh:mm:ss)	00:00:00 (00:00:00-00:49:00)	00:04:06 (00:07:41)
Number of log-ins	18 (3-50)	20.4 (11.4)
Number of days between log-ins	4 (0-19)	4.2 (2.7)
Number of pages visited per log-in	17 (9-36)	17.4 (5.2)
Total number of characters submitted	6533 (346-29,631)	7286 (5093)

Fidelity

At the 3-week point of postrandomization, participants were asked to rate treatment credibility. Treatment credibility was rated highly by the child participants, with a mean score of 6.4 out of 8 (SD 1.5). Furthermore, at the primary endpoint, participants were asked to rate their overall satisfaction with the intervention. Child participants were highly satisfied with the intervention, with a mean score of 24.8 out of 32 (SD 5.2). At the end of the treatment, participants were asked to give their feedback on the intervention within the web-based platform, and they were able to give open-ended responses. Only 59.8% (67/112) of child participants provided feedback. From the conducted quantitative content analysis, four categories emerged relating to implementation fidelity, namely *limitations of ORBIT* (51/67, 76%), which captured how participants felt that overall ORBIT was helpful however was limited by certain factors; *ORBIT as a suitable treatment* (49/67, 73%), which suggested that participants felt that the web-based delivery of treatment for tic disorders was suitable; *problems with using ORBIT* (20/67, 30%), which captured those participants who stated that they felt ORBIT was not helpful to them or was associated with negative factors; and *feeling supported* (19/67, 28%; see Multimedia Appendix 3 for a full list of content analysis categories and codes), where participants mentioned that they felt supported in a way they had never been before (eg, by their therapist). The main code relating to *limitations of ORBIT* centered on *improvement required* (n=33). This code captured anything related to the intervention being unhelpful or inappropriate. Examples include repetitiveness of treatment, the treatment being too short or too long, unhelpful aspects, and suggested improvements. Two child participants reported technical issues with the ORBIT platform, which was related to intermittent problems with connectivity. Despite this, many participants felt the intervention was acceptable as a treatment, with the largest number of participants being coded as *a positive experience of ORBIT* (n=42), which was part of the category *ORBIT as a suitable treatment* and related to being pleased by having taken part and finding it enjoyable while recommending the treatment to other CYP with tic disorders.

Although satisfaction was rated highly, some participants felt that the role of the therapist was somewhat misleading. This was captured by the theme *expectations of role of the therapist*. Some felt that a therapist was not needed for the delivery of the intervention (quote 7). Some parents agreed with the sentiment that they could have completed the therapy without the assistance of a therapist (quote 8).

The term *therapist* itself was felt to be somewhat misjudged as a label, and they viewed the therapist more as a motivator (quote 9). The therapists themselves concurred with this, and perhaps they should not have been called a *therapist* within the ORBIT study (quote 10).

At the end of the interviews, participants were asked if they had any recommendations to improve the intervention, and the overriding majority felt that a mobile application was needed in future iterations of the intervention (quote 11). Some of the older CYP felt that the content and presentation of the intervention were childlike and aimed more toward younger participants. Therefore, they felt there could be two separate versions: one for teenagers and one for young children (quote 12).

Adaptations

Regarding adaptations, the intervention did not appear to evolve in any way from the original plans. Instead, there appeared to be consistency in the way the intervention was delivered and received. Interviews with therapists confirmed that consistency was maintained during delivery. For example, they created a list of standardized responses to common queries (quote 13). Parts of the intervention were designed to be adapted by the user and tailored to their needs and preferences, such as the *tic stopwatch* and *tic ladder* (hierarchy of exposure exercises), which was confirmed in the interviews (quote 14). Another participant adapted the intervention to make it easier to complete by altering some activities to make them more user friendly (quote 15).

Contextual Factors Influencing Intervention Implementation

To establish a measure of intervention implementation that captured both the breadth and depth of participants' usage, a principal component analysis with varimax (orthogonal) rotation was conducted on the seven items related to the dose of intervention received. The analysis suggested a two-factor model. The strongest factor accounted for 47% of the variance (Eigenvalue 3.3; Table 6) and seemed to capture the strength of engagement with the intervention. Factor scores ranged from −2.65 to 2.26 with a mean of 0.001 (SD 0.99), and these scores were used as the engagement measure.

A two-tailed *t* test found that participants enrolled at the London site (mean 0.25, SD 0.90) scored significantly higher on engagement than those enrolled at the Nottingham site (mean −0.22, SD 1.03; t_{109}=−2.58; *P*=.01). Moreover, those who were self-referred (mean 0.16, SD 0.94) scored higher on engagement than those referred by clinics (mean −0.24, SD 1.04; t_{109}=−2.06; *P*=.04). Spearman rho (ρ) correlations were used to determine the association between engagement and various contextual factors. CYP's engagement factor scores were strongly correlated with parents' chapter completion (n=111; ρ=0.73; *P*<.001) and moderately correlated with therapist time for parents (n=111; ρ=0.46; *P*<.001). There were no significant relationships between the CYP's engagement factor score and age, parental education, IMD, TTSS at baseline, or MFQ baseline score. There were also no statistically significant relationships among the child's gender, comorbidities, or use of tic medication and CYP's engagement.

Multiple linear regression was conducted, with CYP's engagement factor score as the dependent variable, and site, child's age, child's gender, IMD, TTSS, referral method, parental education, therapist time for parents, and parents' chapter completion as the independent variables. The results of the simultaneous regression indicated that, collectively, the independent variables had a significant amount of variance in the CYP's engagement factor score ($F_{10,100}$=20.84; *P*<.001; R^2=0.64). There was no evidence of multicollinearity, with all tolerances above 50% and all variance inflation factors below 2. Only parents' chapter completion (β=.69; t_{110}=10.18; *P*<.001) and therapist time for parent (β=.19; t_{110}=2.95; *P*=.004) were significant independent predictors in the model.

The theme of *parental persuasiveness* was generated in the framework category *participant contextual factors*. Many of the parents interviewed outlined that they were often the main motivating force behind their child's level of engagement by reminding their child to practice the learned techniques (quote 16). Some parents found motivating their child to engage very challenging (quote 17), and this was even more challenging for those with children who have comorbidities (quote 18).

Some parents found it difficult to support their children because of hectic schedules, which was captured by the theme of *busy lives* (quote 19). Although under the theme of *high motivation levels*, we found highly engaged CYP without their parents' persuasion (quote 20).

Table 6. Summary of principal component analysis for children's usage data for the Online Remote Behavioral Intervention for Tics intervention (n=111).

Item	Factor loadings	
	Factor 1: engagement	Factor 2: sporadic use
Number of log-ins	0.90	—[a]
Chapters completed	0.79	—
Total therapist time per child	0.76	—
Total number of characters submitted	0.74	—
Number of days between log-ins	−0.63	0.54
Number of pages visited per log-in	−0.41	0.80
Telephone time with the therapist	−0.44	−0.46
Eigenvalue	3.3	1.5
Percentage of variance	47	21

[a] Item not loaded onto factor.

Discussion

Principal Findings

This process evaluation used a mixed methods approach to investigate the extent to which the ORBIT intervention was implemented as planned within the RCT and explore participants' experiences with the intervention and the contextual factors influencing children's engagement. Doing so made it possible to identify reasons for variation in uptake, usage, and engagement, reflect on how implementation may ultimately give us greater confidence in the outcomes, and outline lessons for potential future implementation within routine care. The uptake of the intervention was high, with 88.4% (99/112) of participants receiving the predefined minimum effective dose of the first four chapters completed. The median uptake was eight chapters, and only 1 child failed to access any chapters. Fidelity of delivery was also excellent, with participants reporting high levels of satisfaction and acceptability.

The intended sample of CYP with a diagnosed tic disorder was reached, with 7.1% (8/112) of families residing in the most deprived areas (IMD quintile 1) and more than a quarter (29/112, 25.6%) of the families residing in the least deprived areas (IMD quintile 5). As more than half (60/112, 53.5%) of the CYP's mothers had completed university education, against a UK average of 42% [36], it seems that more advantaged families may have been overrepresented. Perhaps the requirement to have broadband internet access and regular use of a computer with mobile phone text messaging facilities to participate in the study may have differentially impacted participants in the most deprived IMD quintile. This is a concern, as one of the aims of the ORBIT was to increase access to evidence-based therapeutic interventions for CYP with tic disorders. In particular, access to services is generally limited to those from lower economic backgrounds [37]. However, the initial baseline visit with associated travel may have been a disincentive to more disadvantaged families—a limitation that would not be relevant if ORBIT was delivered entirely remotely in routine care rather than as part of an RCT. Moreover, there is no evidence that socioeconomic factors influence CYP's engagement with ORBIT. Furthermore, children's age, tics severity, well-being, and comorbidities did not appear to influence the child's level of engagement with the intervention, providing further evidence that it would have a wide reach within routine clinical care. However, because of the various factors related to this RCT as opposed to routine care, caution should be taken when interpreting the results of this study concerning reach.

The London study site, self-referral, and higher parental engagement were all associated with higher levels of engagement. The London site is a world-renowned center of excellence for pediatric care, which may have increased parents' motivation for treatment. However, the only independent predictor of child engagement in the multivariate analysis was the level of parental engagement with intervention, as measured by their chapter completion and time with the therapist. This is consistent with previous literature [38-41], which found that parental involvement was particularly key for younger CYP to assist with their engagement with therapeutic interventions, which in turn leads to better outcomes [42-44]. It has been shown in the literature that parental engagement may impact a provider's ability to implement parent- and family-focused evidence-based treatment with fidelity [42]. Therefore, it is crucial to understand the role of parental support in the implementation of DHIs for children, as without attention to the key processes of child and family engagement, efforts to improve the effectiveness and efficiency of the treatment are less likely to succeed. Furthermore, it is crucial to assess whether parental support also predicts intervention efficacy and the mechanisms through which its impact is achieved.

An interesting finding was the use and interactions with the therapists in this study. Therapists interacted on the web with their assigned child participants an average of 6 minutes per child each week, which is lower than the 24 minutes average time per week participants interacted with their therapist in the Swedish pilot trial, on which ORBIT is based [14]. However, in the United Kingdom, study therapists were encouraged to use preprepared scripts to respond to the participants. Their responsibilities involved reinforcing the ORBIT treatment material with the aim of spending around 6 minutes a week in response to each child. Detailed analysis of the content of therapists' interactions is outside the remit of this study, but it is apparent from qualitative interviews that many participants felt that the term *therapist* was somewhat misleading. Some participants felt that *therapist* had connotations of a clinically-trained individual delivering an intervention. This may have limited their reliance on therapists. Therefore, in any implementation of this intervention within routine health care, it would be sensible to alter the title to *coach*, *guide*, or *mentor* as this better reflects the therapist's role.

Strengths and Limitations

To the best of our knowledge, this study is one of the first to conduct an in-depth mixed methods process evaluation of a complex intervention aimed at CYP with Tourette syndrome and chronic tic disorders. A number of important findings emerged from the process data, which helped us characterize the intervention's implementation within an RCT and provided lessons for potential future implementation within routine care. However, these lessons can only be fully realized once the main RCT outcome data have been analyzed. Furthermore, a principal component analysis of participants' usage data provided an objective, reliable, and comprehensive measure of engagement to explore the role of contextual factors.

However, this study has some limitations. First, there was the issue of potential recruitment bias. It may have been that the more motivated families self-referred to the trial and that recruitment from clinics was skewed toward punctual, frequent attenders, in contrast to patients with multiple missed appointments. This may have limited the power of this process evaluation to detect socioeconomic biases in engagement. Second, although comprehensive, the information on uptake cannot fully capture the quality and quantity of engagement with the ORBIT. For example, indices such as chapter completion, number of pages visited, and number of log-ins may not fully capture factors such as level of attention to practice exercises. Finally, and perhaps most crucially, a major limitation was that it was not possible to interview those who had not taken part in the RCT or reach those who had withdrawn early from the study. Their perspective is vital to fully understand the factors influencing engagement with DHIs.

Conclusions

We conclude that the intervention had high fidelity of delivery and was evaluated positively by CYP, although some participants suggested some minor improvements, and the nature of the RCT may have constrained reach. Parental engagement was a strong, independent predictor of intervention implementation, which has important implications for designing and implementing digital therapeutic interventions in child and adolescent mental health services.

Acknowledgments

The authors thank Tourettes Action for their ongoing support with the ORBIT trial and particularly acknowledge Dr Seonaid Anderson for her help and advice. The authors also thank Patient and Public Involvement members for their help in revising the interview schedules, including James Bungay, Claire Bungay, Sandra Wang, and Marco Wang. Finally, the authors thank all participants who consented to be interviewed. All research at the Great Ormond Street Hospital National Health Service (NHS) Foundation Trust and University College London Great Ormond Street Institute of Child Health is made possible by the National Institute for Health Research (NIHR) Great Ormond Street Hospital Biomedical Research Center. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. This process evaluation was funded by the NIHR MindTech MedTech Co-operative and NIHR Nottingham Biomedical Research Center Mental Health and Technology Theme. The ORBIT trial is funded by the NIHR Health Technology Assessment (ref 16/19/02). CG, EBD, and CH acknowledge the financial support of the NIHR Nottingham Biomedical Research Center and NIHR MindTech MedTech Co-operative. This research was supported by the NIHR Biomedical Research Center at the Great Ormond Street Hospital for Children's NHS Foundation Trust and University College London.

Authors' Contributions

CG, KK, CLH, and EBD contributed to the mixed methods process evaluation conception and developed the interview guides. KK conducted data collection and initial analysis of both quantitative and qualitative data. KK wrote the initial draft and the subsequent revisions. CG, CLH, and EBD provided a critical review and editing of the initial and subsequent drafts of the manuscript. All authors critically revised the manuscript and read and approved the final manuscript. CH is the principal investigator in the ORBIT trial, designed the trial, and wrote the original grant application.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Analytic framework.
[DOCX File , 19 KB - jmir_v23i6e25470_app1.docx]

Multimedia Appendix 2
Qualitative quotes.
[DOCX File , 17 KB - jmir_v23i6e25470_app2.docx]

Multimedia Appendix 3
Content analysis tables.
[DOCX File , 18 KB - jmir_v23i6e25470_app3.docx]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5. Washington, DC: American Psychiatric Association Publishing; 2013:1-991.
2. Lichter D, Finnegan S. Influence of gender on Tourette syndrome beyond adolescence. Eur Psychiatry 2015 Feb;30(2):334-340. [doi: 10.1016/j.eurpsy.2014.07.003] [Medline: 25193042]
3. Freeman RD, Fast DK, Burd L, Kerbeshian J, Robertson MM, Sandor P. An international perspective on Tourette syndrome: selected findings from 3,500 individuals in 22 countries. Dev Med Child Neurol 2000 Jul;42(7):436-447 [FREE Full text] [doi: 10.1017/s0012162200000839] [Medline: 10972415]
4. Leckman JF, Zhang H, Vitale A, Lahnin F, Lynch K, Bondi C, et al. Course of tic severity in Tourette syndrome: the first two decades. Pediatrics 1998 Jul;102(1 Pt 1):14-19. [doi: 10.1542/peds.102.1.14] [Medline: 9651407]
5. Shprecher D, Kurlan R. The management of tics. Mov Disord 2009 Jan 15;24(1):15-24 [FREE Full text] [doi: 10.1002/mds.22378] [Medline: 19170198]
6. Stroup TS, Gray N. Management of common adverse effects of antipsychotic medications. World Psychiatry 2018 Oct 07;17(3):341-356 [FREE Full text] [doi: 10.1002/wps.20567] [Medline: 30192094]
7. McGuire JF, Piacentini J, Brennan EA, Lewin AB, Murphy TK, Small BJ, et al. A meta-analysis of behavior therapy for Tourette syndrome. J Psychiatr Res 2014 Mar;50:106-112. [doi: 10.1016/j.jpsychires.2013.12.009] [Medline: 24398255]
8. Whittington C, Pennant M, Kendall T, Glazebrook C, Trayner P, Groom M, et al. Practitioner review: treatments for Tourette syndrome in children and young people - a systematic review. J Child Psychol Psychiatry 2016 Sep;57(9):988-1004 [FREE Full text] [doi: 10.1111/jcpp.12556] [Medline: 27132945]
9. Hollis C, Pennant M, Cuenca J, Glazebrook C, Kendall T, Whittington C, et al. Clinical effectiveness and patient perspectives of different treatment strategies for tics in children and adolescents with Tourette syndrome: a systematic review and qualitative analysis. Health Technol Assess 2016 Jan;20(4):1-450, vii-viii [FREE Full text] [doi: 10.3310/hta20040] [Medline: 26786936]
10. Novotny M, Valis M, Klimova B. Tourette syndrome: a mini-review. Front Neurol 2018;9:139 [FREE Full text] [doi: 10.3389/fneur.2018.00139] [Medline: 29593638]
11. Hollis C, Falconer C, Martin J, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503 [FREE Full text] [doi: 10.1111/jcpp.12663] [Medline: 27943285]
12. Ricketts EJ, Goetz AR, Capriotti MR, Bauer CC, Brei NG, Himle MB, et al. A randomized waitlist-controlled pilot trial of voice over internet protocol-delivered behavior therapy for youth with chronic tic disorders. J Telemed Telecare 2016 Apr;22(3):153-162 [FREE Full text] [doi: 10.1177/1357633X15593192] [Medline: 26169350]
13. Himle MB, Freitag M, Walther M, Franklin SA, Ely L, Woods DW. A randomized pilot trial comparing videoconference versus face-to-face delivery of behavior therapy for childhood tic disorders. Behav Res Ther 2012 Sep;50(9):565-570. [doi: 10.1016/j.brat.2012.05.009] [Medline: 22743661]
14. Andrén P, Aspvall K, de la Cruz L, Wiktor P, Romano S, Andersson E, et al. Therapist-guided and parent-guided internet-delivered behaviour therapy for paediatric Tourette's disorder: a pilot randomised controlled trial with long-term follow-up. BMJ Open 2019 Feb 15;9(2):e024685 [FREE Full text] [doi: 10.1136/bmjopen-2018-024685] [Medline: 30772854]
15. Hall C, Davies E, Andrén P, Murphy T, Bennett S, Brown B, ORBIT Trial team.. Investigating a therapist-guided, parent-assisted remote digital behavioural intervention for tics in children and adolescents-'Online Remote Behavioural Intervention for Tics' (ORBIT) trial: protocol of an internal pilot study and single-blind randomised controlled trial. BMJ Open 2019 Jan 03;9(1):e027583 [FREE Full text] [doi: 10.1136/bmjopen-2018-027583] [Medline: 30610027]
16. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]

17. Glasgow R, Lichtenstein E, Marcus A. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health 2003 Aug;93(8):1261-1267 [FREE Full text] [doi: 10.2105/ajph.93.8.1261] [Medline: 12893608]
18. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
19. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Br Med J 2015 Mar 19;350(mar19 6):h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
20. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
21. 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]
22. Hermes E, Lyon A, Schueller S, Glass J. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J Med Internet Res 2019 Jan 25;21(1):e11752 [FREE Full text] [doi: 10.2196/11752] [Medline: 30681966]
23. Hawe P, Shiell A, Riley T. Complex interventions: how "out of control" can a randomised controlled trial be? Br Med J 2004 Jun 26;328(7455):1561-1563 [FREE Full text] [doi: 10.1136/bmj.328.7455.1561] [Medline: 15217878]
24. Bumbarger B, Perkins D. After randomised trials: issues related to dissemination of evidence ‐ based interventions. J Child Serv 2008 Apr 12;3(2):55-64 [FREE Full text] [doi: 10.1108/17466660200800012]
25. Khan K, Hollis C, Hall C, Davies E, Mataix-Cols D, Andrén P, et al. Protocol for the process evaluation of the Online Remote Behavioural Intervention for Tics (ORBIT) randomized controlled trial for children and young people. Trials 2020 Jan 02;21(1):6 [FREE Full text] [doi: 10.1186/s13063-019-3974-3] [Medline: 31898510]
26. English indices of deprivation 2019: technical report. Ministry of Housing, Communities & Local Government, Government of UK. 2019. URL: https://www.gov.uk/government/publications/english-indices-of-deprivation-2019-technical-report [accessed 2021-06-07]
27. Leckman JF, Riddle MA, Hardin MT, Ort SI, Swartz KL, Stevenson J, et al. The Yale Global Tic Severity Scale: initial testing of a clinician-rated scale of tic severity. J Am Acad Child Adolesc Psychiatry 1989 Jul;28(4):566-573. [doi: 10.1097/00004583-198907000-00015] [Medline: 2768151]
28. Costello EJ, Angold A. Scales to assess child and adolescent depression: checklists, screens, and nets. J Am Acad Child Adolesc Psychiatry 1988 Nov;27(6):726-737. [doi: 10.1097/00004583-198811000-00011] [Medline: 3058677]
29. Daviss WB, Birmaher B, Melhem N, Axelson D, Michaels S, Brent D. Criterion validity of the Mood and Feelings Questionnaire for depressive episodes in clinic and non-clinic subjects. J Child Psychol Psychiatry 2006 Sep;47(9):927-934 [FREE Full text] [doi: 10.1111/j.1469-7610.2006.01646.x] [Medline: 16930387]
30. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep 6;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
31. Fusch P, Ness L. Are we there yet? Data saturation in qualitative research. Qual Rep 2015 Sep 8:A [FREE Full text] [doi: 10.46743/2160-3715/2015.2281]
32. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess B, editors. Analyzing Qualitative Data. London: Routledge; 1994:173-194.
33. Gale N, 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]
34. Hsieh H, Shannon S. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288 [FREE Full text] [doi: 10.1177/1049732305276687] [Medline: 16204405]
35. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. Br Med J 2010 Sep 17;341:c4587. [doi: 10.1136/bmj.c4587] [Medline: 20851841]
36. Graduates in the UK labour market. Office for National Statistics. 2017. URL: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/graduatesintheuklabourmarket/2017#steady-increase-in-the-number-of-graduates-in-the-uk-over-the-past-decade [accessed 2021-06-07]
37. Packness A, Waldorff F, Christensen RD, Hastrup L, Simonsen E, Vestergaard M, et al. Impact of socioeconomic position and distance on mental health care utilization: a nationwide Danish follow-up study. Soc Psychiatry Psychiatr Epidemiol 2017 Nov;52(11):1405-1413 [FREE Full text] [doi: 10.1007/s00127-017-1437-2] [Medline: 28849245]
38. Cobham VE. Do anxiety-disordered children need to come into the clinic for efficacious treatment? J Consult Clin Psychol 2012 Jun;80(3):465-476. [doi: 10.1037/a0028205] [Medline: 22545740]
39. Lyneham HJ, Rapee RM. Evaluation of therapist-supported parent-implemented CBT for anxiety disorders in rural children. Behav Res Ther 2006 Sep;44(9):1287-1300. [doi: 10.1016/j.brat.2005.09.009] [Medline: 16313883]

40. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, et al. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
41. Thirlwall K, Cooper P, Karalus J, Voysey M, Willetts L, Creswell C. Treatment of child anxiety disorders via guided parent-delivered cognitive-behavioural therapy: randomised controlled trial. Br J Psychiatry 2013 Dec;203(6):436-444 [FREE Full text] [doi: 10.1192/bjp.bp.113.126698] [Medline: 23969483]
42. Haine-Schlagel R, Walsh N. A review of parent participation engagement in child and family mental health treatment. Clin Child Fam Psychol Rev 2015 Jun;18(2):133-150 [FREE Full text] [doi: 10.1007/s10567-015-0182-x] [Medline: 25726421]
43. Nock M, Ferriter C. Parent management of attendance and adherence in child and adolescent therapy: a conceptual and empirical review. Clin Child Fam Psychol Rev 2005 Jun;8(2):149-166 [FREE Full text] [doi: 10.1007/s10567-005-4753-0] [Medline: 15984084]
44. Vernon T, Koegel R, Dauterman H, Stolen K. An early social engagement intervention for young children with autism and their parents. J Autism Dev Disord 2012 Dec;42(12):2702-2717 [FREE Full text] [doi: 10.1007/s10803-012-1535-7] [Medline: 22527708]

Abbreviations

CYP: children and young people
DHI: digital health intervention
IMD: index of multiple deprivation
MFQ: Mood and Feelings Questionnaire
MRC: Medical Research Council
NHS: National Health Service
NIHR: National Institute for Health Research
ORBIT: Online Remote Behavioral Intervention for Tics
RCT: randomized controlled trial
TTSS: Total Tic Severity Score
YGTSS: Yale Global Tic Severity Scale

Edited by R Kukafka; submitted 04.11.20; peer-reviewed by J van de Griendt, H McCall; comments to author 23.12.20; revised version received 15.01.21; accepted 16.03.21; published 21.06.21.

Please cite as:
Khan K, Hollis C, Hall CL, Murray E, Davies EB, Andrén P, Mataix-Cols D, Murphy T, Glazebrook C
Fidelity of Delivery and Contextual Factors Influencing Children's Level of Engagement: Process Evaluation of the Online Remote Behavioral Intervention for Tics Trial
J Med Internet Res 2021;23(6):e25470
URL: https://www.jmir.org/2021/6/e25470
doi:10.2196/25470
PMID:34152270

Original Paper

Facebook Users' Interactions, Organic Reach, and Engagement in a Smoking Cessation Intervention: Content Analysis

Dávid Pócs[1], MD; Otília Adamovits[1]; Jezdancher Watti[1], MD; Róbert Kovács[1], MD; Oguz Kelemen[1], MD, PhD

[1] Department of Behavioral Sciences, Faculty of Medicine, University of Szeged, Szeged, Hungary

Corresponding Author:
Dávid Pócs, MD
Department of Behavioral Sciences
Faculty of Medicine
University of Szeged
Szentháromság u. 5
Szeged, 6722
Hungary
Phone: 36 62 420 530
Fax: 36 62 545 968
Email: drpocsdavid@gmail.com

Abstract

Background: Facebook can be a suitable platform for public health interventions. Facebook users can express their reaction to the given social media content in many ways using interaction buttons. The analysis of these interactions can be advantageous in increasing reach and engagement of public health interventions.

Objective: This research aimed at understanding how Facebook users' interactions correlate with organic reach and engagement regarding the same smoking cessation support contents.

Methods: The study population consisted of Facebook users who were reached by a public smoking cessation support page without advertising. We included 1025 nonpaid Facebook posts (N=1025) which used smoking cessation strategies based on a motivational interviewing counseling style. The following data were collected from the "Post Details": the number of people who saw the given nonpaid content (organic reach) which consisted of fan and nonfan reach according to previous "page like" activity; each rate of "engagement indicators" (such as the symbols of "like," "love," "haha," "wow," "sad," "angry"; or other interactions: "shares," "comments," "clicks"); and the rate of negative Facebook interactions (eg, "post hides" or "unlike of page"). Overall, these data were analyzed with the Spearman correlation method.

Results: Surprisingly, we found a significant negative correlation between organic reach and the "like" reaction (r_s=–0.418; *P*<.001). The strongest significant positive correlations of organic reach were observed with the "haha" reaction (r_s=0.396; *P*<.001), comments (r_s=0.368; *P*<.001), and the "love" reaction (r_s=0.264; *P*<.001). Furthermore, nonfan reach correlated positively with "shares" (r_s=0.388; *P*<.001) and clicks (r_s=0.135; *P*<.001), while fan reach correlated positively with the "haha" reaction (r_s=0.457; *P*<.001), comments (r_s=0.393; *P*<.001), and the "love" reaction (r_s=0.310; *P*<.001). Contrary to expectations, the "like" reaction was sharply separated by significant negative correlations from "wow" (r_s=–0.077; *P*=.013), "sad" (r_s=–0.120; *P*<.001), "angry" reactions (r_s=–0.136; *P*<.001), and comments (r_s=–0.130; *P*<.001). Additionally, a high rate of negative Facebook interactions was significantly associated with "wow" (r_s=0.076; *P*=.016) and "sad" reactions (r_s=0.091; *P*=.003).

Conclusions: This study has shown that it is possible to hypothesize a disadvantage of the "like" reaction and advantages of other interactions (eg, the "haha" reaction or "comments") in content algorithmic ranking on Facebook. In addition, the correlational analysis revealed a need of a further categorization to fan-specific interactions (eg, "haha" or "love" reactions) and nonfan-specific interactions (eg, "shares" and "clicks"). Regarding the direction of the correlations, these findings suggest that some interactions (eg, negative Facebook interactions, "wow," "sad," and "angry" reactions) may decrease the engagement, while other interactions ("like," "love," "haha" reactions, "shares," and "clicks") may increase the engagement during Facebook-based smoking cessation interventions. This hypothesis-generating research offers an important insight into the relationship between organic reach, engagement, and Facebook users' interactions for public health professionals who design Facebook-based interventions.

(J Med Internet Res 2021;23(6):e27853) doi:10.2196/27853

KEYWORDS

smoking; smoking cessation; behavior; health behavior; internet; social media; love; comment; motivation; language; public health

Introduction

Reaching to Facebook Users

Facebook is a widely available social media platform, which could be highly relevant for people who seek help with health behavior change [1]. Facebook is also used by public health organizations to communicate health messages [2,3], especially for reducing smoking [4-7]. Social media contents which support smoking cessation can be more cost-effective than television advertising [8]. Facebook can also be a useful tool to contact hard-to-reach smokers [9]. On this platform a major intention of smoking cessation or other public health interventions could be to reach a large number of users through social media contents [10]. A commonly used measure of dissemination can be the Facebook "post reach" data, which represent the number of people who saw the given social media content [11,12]. Facebook provides access to these reach data and allows page administrators to increase the post reach by paying [13]. Therefore, "paid post reach" and "organic post reach" data must be distinguished. These data refer to the number of people who saw a paid or a nonpaid social media content, respectively [11,14]. Previous research has found that nonpaid (organic) reach is associated with higher engagement than paid reach [11]. "Post reach" data can also be divided into fan reach and nonfan reach based on the previous usage of the "page like" button.

An increasing number of contents are being published on Facebook; however, the number of posts a Facebook user is able to see at a time is limited. Therefore, Facebook must filter the social media contents for the users, and contents compete with each other to reach Facebook users [15,16]. This phenomenon has opened an exciting direction for research to reveal the opportunities to increase organic post reach in Facebook-based public health interventions [17,18]. Because of the increasing number of posts, Facebook provides only the most relevant content to each user [15,18]. The basis of this highly personalized filtering is the "Facebook algorithm of content ranking," which ranks all available posts that can be displayed on a user's News Feed. This algorithm is reviewed annually, but the details of the Facebook algorithm are unknown (ie, not published) [11,15]. However, some major elements of the algorithm which may determine the Facebook News Feed content are suspected, such as the Facebook user's past activity (eg, sending a message to the page or using the "page like" button); the past activity of the Facebook page (eg, violating the Facebook Community Standards); the performance of the given post (the rate of "like," "shares," "clicks," or other interactions); the post type (eg, image or video); or the timing of the published content (eg, novelty) [2,11,16-18].

These indicate a need to understand how to correlate Facebook users' interactions with organic (fan and nonfan) reach during a smoking cessation intervention, which is a major aim of our research. The platform of the investigated smoking cessation intervention was a public Facebook page, which published social media contents based on a motivational interviewing counseling style. We have summarized our problem statements, aims, and research questions regarding organic reach in Textbox 1. The relationship between total organic reach and different Facebook interactions on a post level may highlight the way in which the Facebook algorithm ranks available contents according to Facebook users' interactions. For example, some interactions can have a higher impact on content algorithmic ranking due to stronger positive correlation with organic reach, while other interactions can have a negative role in content ranking because of negative correlation with organic reach. Likewise, the association between fan reach, nonfan reach, and Facebook interactions may show which element is preferred by the algorithm: for example, Facebook users' interactions on the given post or previous "page like" activity. For instance, some interactions, which correlate negatively with fan reach, can have a higher impact on content ranking than previous "page like." Finally, the correlational analysis between organic reach and Facebook interactions separately for each year may also show how the Facebook algorithm of content ranking is modified annually. This study is an exploratory research, which analyses a set of data searching for correlations, and then proposes hypotheses which may then be tested in subsequent studies.

Interactions on Facebook

A Facebook interaction on a post level is defined as any action on specific buttons that the user performs in relation to content, and it can be divided into 3 groups: "positive interactions" (reactions, shares, comments), "neutral interactions" (clicks), and "negative interactions" (post hides, hides of all posts, reports of spam, unlike of page) [2,11,19-22]. The Facebook reactions (such as like, love, haha, wow, sad, angry) are designed to give users a more nuanced way to express their emotions, and these could have a special role in public health interventions [22]. For example, a tobacco cessation study has shown that receiving 1 "like" reaction on the Facebook-based intervention platform is associated with smoking reduction by approximately 1 cigarette per week [23]. The "like" reaction was introduced in 2009, while other reactions were instituted in 2016, so there has been little quantitative analysis of "love," "haha," "wow," "sad," or "angry" reactions in public health interventions so far [2,22].

The "share" interaction covers several ways of sending the content with optional privacy settings to others, such as share content on the user's or a friend's timeline, share in a Facebook group, share to a Facebook page, or send as a message via Facebook [24,25]. Internet users tend to share content when they perceive a low risk of reputational damage and high benefits to their social network; nevertheless, further investigation is needed to find what makes public health content be shared [11,26]. With the "comment" button, users can publish a text or an image message under a given Facebook post [26,27]. The total number of clicks contains more neutral actions on the content, such as selecting a website, viewing the Facebook page profile, or expanding photos to full screen [28].

Facebook users' interactions can be placed in a theoretical framework of "engagement" [29]. In previous behavioral science studies, "engagement" of digital behavior change interventions was interpreted as the extent of usage [30,31]. In Facebook-based interventions, engagement can be defined as a group of Facebook interactions. These "engagement indicators" include "reactions," "shares," "comments," and "clicks" [11,12,32]. Several attempts have been made to determine the "depth" or the "spectrum" of engagement on the basis of the fact that some Facebook interactions (eg, "shares") can have higher engagement than others (eg, the "like" reaction) [33]. However, little is known about the spectrum of engagement, and it is still not clear which Facebook interactions may stand at a similar engagement level, or which interactions may occur as a combination of engagement indicators [29,33]. Furthermore, the spectrum of engagement raises a need to understand the relationship between positive emotional interactions (eg, "like" or "love" reactions) and negative emotional interactions (eg, "sad" or "angry" reactions). As opposed to engagement, negative Facebook interactions (post hides, hides of all posts, reports of spam, unlike of page) could indicate a neglect of the Facebook-based intervention as a negative engagement indicator. Nevertheless, there has been little discussion in the literature about the role of negative Facebook interactions in online public health interventions. Textbox 1 provides the problem statements, aims, and research questions related to engagement.

Textbox 1. Problem statements, aims, and research questions of this study.

Problem Statement 1

Broad reach is an essential condition for a successful web-based public health intervention. Organic reach on the Facebook platform depends on a hidden algorithm, which includes Facebook users' interactions.

Aim

To understand the association between organic reach and Facebook interactions during a smoking cessation intervention.

Research Questions

- What is the relationship between total organic reach and Facebook interactions on a post level during a smoking cessation intervention?
- How fan reach and nonfan reach of identical intervention contents are influenced by different Facebook interactions?
- How do the correlations between total organic reach and Facebook interactions change each year (in parallel with a modified Facebook algorithm)?

Problem Statement 2

High engagement rate is a relevant feature for a successful web-based public health intervention. Engagement on Facebook can be defined as a combination of interactions. However, some Facebook interactions may express positive emotions, while others may express negative emotions.

Aim

To assess the correlations between positive and negative emotional interactions during a Facebook-based smoking cessation intervention.

Research Questions

- How does the direction of correlations change between engagement indicators which express positive and negative emotion during a smoking cessation intervention?
- How does the strength of positive correlations change between engagement indicators at similar smoking cessation support contents?
- What is the relationship between engagement indicators and negative Facebook interactions on a post level during a smoking cessation intervention?

Methods

Stimuli

The investigated Hungarian "Cigarette break" Facebook page [34] was a nonbusiness and nongovernmental smoking cessation intervention. The primary aim of this program was to avoid frightening and judgmental communication about smoking, and support smoking cessation using a motivational interviewing counseling style (ie, building on collaboration, partnership, and empathetic understanding; emphasizing the autonomy of smokers; supporting self-efficacy). Secondary aims of the intervention were harm reduction in smokers, relapse prevention in former smokers, and developing assertive and social support skills in nonsmokers. Motivational interviewing strategies have been deliberately involved in the moderator work and the creation of Facebook posts. Social media contents were usually published daily at 5 pm on weekdays and at 2 pm on weekends from the beginning (March 7, 2017). The page was edited and managed by the authors, university students, and health care professionals experienced in motivational interviewing.

In all, 1269 social media contents were made during the research period, between March 7, 2017 and August 14, 2020. We excluded 244 Facebook posts in accordance with the following exclusion criteria. First, we excluded 60 "boosted" Facebook posts, which were promoted by paid Facebook advertising after publication to increase reach and engagement. Boosted posts can influence the results through the rate of interactions [11,33]. They can reach more Facebook users who are active in engagement before the advertising, rather than other passive users. Second, 24 video posts were also excluded because the evaluation of the interactions was fundamentally different from the assessment of the image posts (eg, minutes viewed, 10-second views), and also because the Facebook algorithm evaluates these posts differently during content ranking [2,16].

Third, we excluded 69 administrator's posts and 16 posts which were targeted at nonsmokers. We only included contents that directly addressed to smokers. Fourth, 68 social media contents were excluded because of noncessation topic (eg, second-hand smoking or harm reduction). Only smoking cessation support Facebook posts were included to focus on public health interventions. Finally, we excluded 7 motivational interviewing nonadherent posts, which did not conform perfectly to the spirit of motivational interviewing during the retrospective analysis. It should be emphasized that only original Facebook posts were analyzed to evaluate the users' response given to the same stimulus. Therefore, shared Facebook posts were ignored, because in these cases, Facebook users' responses could have been influenced by other stimuli (eg, the Facebook profile of the person who shared the content), which may have resulted in a higher rate of interactions than the original content [18].

After content exclusion, 1025 original posts were included, which all followed the spirit of motivational interviewing, supported smoking cessation, were targeted at smokers, and did not use specific advertising. The majority of the social media contents included (994/1025, 96.97) were image posts, but some (31/1025, 3.02%) had only texts and links to other public health websites. The stimuli were smoking cessation support contents based on motivational interviewing on a public Facebook page; therefore, the content analysis can only be interpreted in this context. We present some examples of motivational interviewing style contents (stimuli) in Multimedia Appendix 1 to illustrate generalizability of the findings to other Facebook-based smoking cessation interventions. We also show some examples of excluded contents in Multimedia Appendix 2.

It should be highlighted that these social media contents were public, and Facebook users used different interaction opportunities voluntarily and freely, without external coercion. It should be also noted that our researcher identity is transparent to anyone in the description of the Facebook page. In addition, we provided information about the research and results through our publicly available posts.

Participants

The theoretical population of this research was formed by any Facebook user who was reached by the included, nonpaid contents of the investigated public Facebook page. It is difficult to precisely determine the study population because the Facebook platform and the audience of the public Facebook page could have changed during this long research period. Therefore, we used a convenience sample. However, Facebook users' data on a page level were exported from "Facebook Insights" on August 14, 2020, which contains epidemiological characteristics.

It should be emphasized that by creating Facebook profiles, users accept the terms and conditions of Facebook. These terms indicate that their data may be accessed by a third party. With this informed consent, Facebook provides anonymized and aggregate data to the page administrators.

At that time, the investigated Facebook page had 10,098 "Facebook fans," who expressed their interest in and support for the page by a "Facebook page like." Based on the age and gender information they provided in their user profiles, 52.91% of the Facebook page fans (5343/10,098) were women, and 46.97% (4744/10,098) were men, and 83.4% (8457/10,098) were between the ages of 18 and 34 years. The vast majority (9634/10,098, 95.40%) were from Hungary; but some German (96/10,098, 0.5%), Serbian (79/10,098, 0.78%), and Romanian (78/10,098, 0.77%) Facebook locations also presented; 9720/10,098 (96.25%) Facebook fans spoke Hungarian, 275/10,098 (2.72%) English, and 103/10,098 (1.02) other languages, based on the default language setting selected.

By contrast, the monthly total reach of the Facebook page was wider: 96,654 people saw the Facebook page's contents in the last month; 43% of them were women and 57% were men; and 85% were between the ages of 18 and 34 years. The distribution of location and language data was roughly similar to the Facebook fans' characteristics. The monthly total reach consisted of paid reach (66,463 people) and organic reach (30,191 people). Unfortunately, Facebook does not register users' smoking status, therefore the smoking habits of the target population were unknown. However, we surveyed smoking status among the audience of this Facebook page in our previous study, where we found 65% were current smokers, 12% former smokers, and 23% nonsmokers [35]. The majority of the smokers (94%) used tobacco daily, while the minority (6%) used tobacco occasionally [35]. The most common nicotine product used in the sample was cigarette (98%), followed by e-cigarette (32%), hookah (20%), cigar (6%), snus (2%), pipe (2%), and snuff (2%) [35].

Design

Our research method was a hypothesis-generating, retrospective, quantitative content analysis. The Facebook posts' organic reach and interaction data were analyzed with the Spearman correlation method. These data were collected from the "Post Details" and belonged to the same social media content (same stimulus). Therefore, we analyzed data on a post level.

It should be noted that in our research we analyzed only 2 well-known elements of the Facebook algorithmic content ranking: users' interactions on a post level (performance of the given post) and previous "page like" through fan reach (Facebook user's past activity). There was not a remarkable difference in other known elements of the Facebook algorithm: mostly image contents were included (post type), all contents were regularly published at the same time (timing of published content), and the management of the investigated Facebook page was not changed notably (past activity of the Facebook page).

The definitions of organic reach and the different types of interactions were discussed in the "Introduction" section and we summarize them in Textbox 2. Facebook interactions can indicate how social media contents increase the usage of a Facebook-based intervention (reactions, shares, comments, clicks) or decrease it (negative Facebook interactions). We used the total number of negative Facebook interactions during the analysis, because it was available together and not separate in "Post Details." It should be noted that the combination of interactions (eg, the "like" reaction and the "share" interaction) could arrive from different users or the same Facebook user, because these data were summarized. However, the combination of reactions on the same content (eg, “like” and “love” reactions) indicates different Facebook users, because 1 Facebook user could choose only 1 reaction. It should be emphasized again that organic reach and Facebook interaction data were anonymized and aggregated, so Facebook users could not be identified.

Facebook uses a private algorithm for highly personalized filtering of social media contents, which influences the methodology of our research at 2 points. First, the correlation between organic reach and interactions were also analyzed separately by years because Facebook changes this algorithm annually. Second, interaction data had to be corrected for the organic reach: the number of each interaction was divided by the number of people who saw the nonpaid post (organic reach) for the statistical analysis. This correction was necessary because increased organic reach can directly enhance other interaction numbers. In other words, if more Facebook users see the post, they are more likely to use interaction buttons. Consequently, the correlational analysis between organic reach and the total number of Facebook interactions would highlight a simple relationship rather than the impact of the Facebook algorithmic content ranking. However, we used an “interaction rate” to express the frequency of the given interaction at the same organic reach. Facebook uses the same correction of interaction data (called “engagement rate”), which is the number of people who liked, commented, shared, or clicked on the post divided by organic reach. “Engagement rate” can be accessed by page managers in the “Facebook Insights,” and the Facebook algorithm of content ranking may use this rate (as a performance indicator of the given post). Therefore, the “interaction rate” which was used in this study can be advantageous for the correlational analysis between organic reach and Facebook interactions or between Facebook interactions.

Textbox 2. Definitions of organic reach, different interactions, and engagement used in the current research.

Organic Reach

The number of people who saw the given nonpaid social media content.

- *Fan reach*: The number of people who had liked the Facebook page before they saw the given nonpaid social media content.
- *Nonfan reach*: The number of people who had not liked the Facebook page before they saw the given nonpaid social media content.

Facebook Interactions

Any action on specific buttons that the user performs in relation to content.

- *Reactions*: The number of people who used a “like,” “love,” “haha,” “wow,” “sad,” or “angry” reaction button under a given social media content to express their emotions.
- *Share*: The number of people who used the “share” button under a given social media content to send the content with optional privacy settings to others.
- *Comment*: The number of people who used the “comment” button under a given social media content to publish a text or an image message.
- *Click*: The number of people who used any other actions on a given social media content, for example, to select a website, to view the Facebook page profile, or to expand photos to full screen.
- *Negative Facebook interactions*: The total number of people who used the following functions: post hides, hides of all posts, reports of spam, unlike of page.

Engagement

A group of the following Facebook interactions: reactions, share, comment, and clicks. These interactions are called *engagement indicators* and have been defined previously.

Procedure

Our retrospective research aim was to analyze the data of Facebook posts published between March 7, 2017, and August 14, 2020. Two weeks were provided between August 14 and August 28 for the internet users to give a response to the last investigated Facebook post. The data collection was performed between August 28, 2020, and August 30, 2020. First, 1025 Facebook posts were included based on our exclusion criteria (N=1025). Second, organic reach and Facebook interaction data were collected from the “Post Details.” Third, we divided interaction data with organic reach to achieve “interaction rate.” Then, we assessed Spearman correlation between organic reach and Facebook interactions in the research period. In a second analysis, this correlation was investigated separately for each year to evaluate the potential effect of the Facebook algorithm which is modified annually. Finally, we also used Spearman correlation between Facebook interactions.

All analyses were conducted using the SPSS software (IBM). Correlation statistics were performed rather than regression because the cause-and-effect relationship between the investigated variables was unclear. We used the nonparametric Spearman correlation, rather than the Pearson correlation, because of the non-normal distribution of the data. In all analyses, the conditions of the Spearman correlation were met; variables were measured on an interval or ratio scale, variables represented paired observations, and monotonic relationship was detected between the variables using scatterplot test. The *P* value of less than .05 was taken to indicate a significant effect, and the *P* value of less than .001 was taken to indicate a highly significant effect. Original data supporting the results presented in the paper can be found in Multimedia Appendix 3. The data on statistical analyses are available from the corresponding author upon request.

Results

Trends in Reach and Interactions

We summarized the mean and SD of organic reach and Facebook interactions in Table 1. Organic reach is divided into fan and nonfan columns to illustrate the distribution of the average (total) organic reach between these 2 indicators. The average number of organic reach in the research period was 1328 (SD 981), of which 59% were Facebook page fan (783 Facebook users) and 41% were nonfan participants (545 Facebook users). Table 1 shows that the organic reach of social media contents gradually increased from one year to another in parallel with a growing fan reach percentage.

In the research period, the average number of Facebook interactions for 1 social media content was 15.8 "like" reaction (SD 11.7), 0.9 "love" reaction (SD 4.3), 3.2 "haha" reaction (SD 7.3), 0.3 "wow" reaction (SD 2.2), 0.2 "sad" reaction (SD 0.8), 0.1 "angry" reaction (SD 1.3), 2.4 "share" interaction (SD 3.7), 3.2 "comment" interaction (SD 7.1), 84.9 "click" interaction (SD 163.8), and 0.1 negative interaction (SD 0.4). It should be highlighted that the low number of "sad," "angry" reactions, and negative interactions can be attributed to the spirit of motivational interviewing (eg, partnership or acceptance), which was a primary consideration in the creation of the investigated Facebook posts. Table 1 represents that some interactions ("haha" and "love" reactions or comments) followed the notable increase in organic reach from year to year. By contrast, the mean of some interactions (shares or negative Facebook interactions) did not change remarkably annually. Therefore, despite the greater number of Facebook users who saw the social media content, the activity of Facebook users regarding shares and negative Facebook interactions probably decreased.

Table 1. Mean and SD of organic reach and Facebook interactions in the research period together and separately for each year.

Period	Organic reach	Facebook interactions: Engagement indicators, mean (SD)									
		Reaction						Share	Comment	Click	Negative interactions
		Like	Love	Haha	Wow	Sad	Angry				
ALL[a] (N=1025)[b]	1328.30 (981.23)	15.84 (11.7)	0.91 (4.29)	3.17 (7.33)	0.31 (2.20)	0.18 (0.84)	0.11 (1.29)	2.39 (3.68)	3.17 (7.06)	84.89 (163.82)	0.14 (0.40)
2017 (N=215)[c]	709.38 (565.02)	11.78 (6.90)	0.15 (0.99)	0.48 (1.41)	0.08 (0.32)	0.10 (0.63)	0.03 (0.29)	1.67 (2.64)	1.31 (3.40)	70.36 (89.93)	0.14 (0.39)
2018 (N=378)[d]	1380.96 (1107.79)	16.39 (11.4)	0.39 (1.64)	2.73 (5.27)	0.17 (1.98)	0.10 (0.45)	0.10 (0.65)	3.07 (5.00)	3.14 (6.44)	92.30 (216.93)	0.08 (0.32)
2019 (N=299)[e]	1518.32 (775.13)	17.92 (14.2)	1.08 (4.27)	4.58 (8.58)	0.43 (2.72)	0.25 (1.17)	0.07 (0.50)	2.31 (2.48)	3.65 (7.32)	73.31 (97.33)	0.19 (0.46)
2020 (N=133)[f]	1752.00 (1093.45)	16.19 (10.7)	3.22 (9.23)	5.58 (12.08)	0.75 (3.01)	0.35 (1.05)	0.36 (3.31)	1.76 (2.45)	5.17 (10.90)	113.36 (194.08)	0.17 (0.44)

[a] All contents in the research period (2017-2020).

[b] Fan and nonfan reach: 59% and 41%, respectively.

[c] Fan and nonfan reach: 38% and 62%, respectively.

[d] Fan and nonfan reach: 55% and 45%, respectively.

[e] Fan and nonfan reach: 67% and 33%, respectively.

[f] Fan and nonfan reach: 65% and 35%, respectively.

Correlations Between Reach and Interactions

The results of the correlational analysis between organic reach and Facebook interactions are summarized in Table 2. First, we analyzed the organic reach for the research period. The first question in this research area was, "What is the relationship between total organic reach and Facebook interactions?" We found that most of the interactions had positive correlations with total organic reach, except the "like" reaction and the "share" interaction. The most surprising result was a highly significant negative correlation between the "like" reaction and organic reach (*P*<.001). Furthermore, we could not find a significant difference between the "share" interaction and total organic reach in the research period, although the correlational analysis was nearly significant (*P*=.057). In addition, the highly significant positive correlation between negative interactions and organic reach was interesting (*P*<.001). The strongest highly significant positive correlations were observed between total organic reach and the "haha" reaction (r_s=0.396, *P*<.001), comments (r_s=0.368, *P*<.001), and the "love" reaction (r_s=0.264, *P*<.001). In summary, these results suggest that the Facebook algorithm of content ranking took the interactions into account to varying degrees for the calculation of total organic reach.

The second research question was, "How are fan reach and nonfan reach of the investigated contents influenced by different Facebook interactions?" The direction of significant fan reach and nonfan reach correlations was the same for most interactions, except for the "share" and "click" interactions. We found an opposite direction of 2 highly significant correlations between the “share” interaction and fan reach (r_s=–0.177; *P*<.001), and between the “share” interaction and nonfan reach (r_s=0.388; *P*<.001). This result may explain the absence of a significant relationship between the “share” interaction and total organic reach. Furthermore, a highly significant positive correlation was found between “clicks” and nonfan reach (*P*<.001), while there was no significant relationship between “clicks” and fan reach (*P*=.87). Therefore, social media contents which generated a high rate of “shares” and “clicks” reached relatively more nonfan Facebook users than fan Facebook users. By contrast, the strongest highly significant positive correlations were observed between fan reach and the “haha” reaction (r_s=0.457, *P*<.001), comments (r_s=0.393, *P*<.001), and the “love” reaction (r_s=0.310, *P*<.001). By contrast, less strong correlations were found between nonfan reach and the “haha” reaction (r_s=0.153, *P*<.001), comments (r_s=0.192, *P*<.001), and the “love” reaction (r_s=0.135, *P*<001). These results suggest that social media contents which evoked a high rate of “haha” and “love” reactions and “comments” reached relatively more fan Facebook users than nonfan Facebook users.

Second, the Spearman correlation between organic reach and Facebook interactions was tested separately for each year in the research period. The last question in this research area was, “What are the correlations between organic reach and Facebook interactions each year?” The most striking result to emerge from the data is that the negative correlation between total organic reach and the “like” reaction gradually weakened, while positive correlations between total organic reach and some interactions (“haha,” “wow,” “sad,” “angry” reactions, comments, clicks) strengthened year by year. Similar observations can be made for the fan reach and nonfan reach. Taken together, these results might reflect the modifications to the Facebook algorithmic content ranking during the research period, if we assume that the other elements of the Facebook algorithm did not change significantly in those years.

Table 2. Spearman correlation between Facebook interactions and organic reach in the research period together and separately for each year.

Organic reach	Facebook interactions: Engagement indicators (Spearman correlation coefficients)									
	Reaction						Share	Comment	Click	Negative interactions
	Like	Love	Haha	Wow	Sad	Angry				
ALL[a] (N=1025)										
Total	–0.418[b]	0.264[b]	0.396[b]	0.164[b]	0.231[b]	0.160[b]	0.059	0.368[b]	0.076[c]	0.124[b]
Fan	–0.352[b]	0.310[b]	0.457[b]	0.169[b]	0.245[b]	0.173[b]	–0.174[b]	0.393[b]	–0.005	0.089[c]
Nonfan	–0.332[b]	0.135[b]	0.153[b]	0.096[c]	0.144[b]	0.103[c]	0.388[b]	0.192[b]	0.135[b]	0.104[c]
2017 (N=215)										
Total	–0.633[b]	0.028	0.142[c]	0.092	0.113	0.091	0.255[b]	0.266[b]	0.124	0.240[b]
Fan	–0.412[b]	–0.013	0.144[c]	0.086	0.166[c]	0.063	–0.186[c]	0.234[c]	0.147[c]	0.144[c]
Nonfan	–0.549[b]	0.056	0.011	0.071	0.033	0.069	0.392[b]	0.195[c]	0.092	0.196[c]
2018 (N=378)										
Total	–0.296[b]	0.178[c]	0.295[b]	0.137[c]	0.180[b]	0.152[c]	0.157[c]	0.403[b]	0.305[b]	0.013
Fan	–0.202[b]	0.173[c]	0.420[b]	0.140[c]	0.236[b]	0.210[b]	–0.235[b]	0.460[b]	0.154[c]	–0.024
Nonfan	–0.278[b]	0.115[c]	–0.005	0.084	0.060	0.062	0.537[b]	0.171[c]	0.321[b]	0.019
2019 (N=299)										
Total	–0.043	0.301[b]	0.468[b]	0.183[c]	0.286[b]	0.162[c]	0.136[c]	0.358[b]	0.407[b]	0.143[c]
Fan	–0.010	0.321[b]	0.516[b]	0.221[b]	0.251[b]	0.151[c]	–0.047	0.423[b]	0.473[b]	0.120[c]
Nonfan	–0.106	0.174[c]	0.210[b]	0.078	0.248[b]	0.125[c]	0.420[b]	0.119[c]	0.166[c]	0.119[c]
2020 (N=133)										
Total	–0.121	–0.122	0.438[b]	0.206[c]	0.280[c]	0.254[c]	0.258[c]	0.509[b]	0.658[b]	0.289[c]
Fan	–0.089	–0.042	0.428[b]	0.218[c]	0.325[b]	0.259[c]	0.126	0.547[b]	0.717[b]	0.315[b]
Nonfan	–0.112	–0.232[c]	0.326[b]	0.152	0.144	0.164	0.414[b]	0.296[c]	0.383[b]	0.159

[a] Organic reach of all contents in the research period (2017-2020).

[b] Highly significant, *P*<.001 (2-tailed).

[c] Significant, *P*<.05 (2-tailed).

Correlations Between Interactions

We investigated the 1025 social media contents in relation to the Facebook interactions. Table 3 provides the intercorrelations among the 10 types of Facebook interactions. First, the engagement indicators (reactions, shares, comments, clicks) were analyzed with the Spearman correlation method. The first question in this research area was, "How does the direction of correlations change between engagement indicators which express positive and negative emotion?" We found that most of the correlations between engagement indicators had a positive direction; however, negative correlations were observed in some cases, which might be associated with negative emotional characteristics (eg, "sad" and "angry" reactions). Highly significant negative correlations were found between "sad" and "like" reactions (*P*<.001) as well as between "angry" and "like" reactions (*P*<.001). The significant negative correlation between "wow" and "like" reactions (*P*=.013) as well as between "comment" and "like" interactions (*P*<.001) may suggest that Facebook users could express their negative emotions also with the "wow" reaction or a negative emotional comment. By contrast, the significant negative correlations between "comment" and "share" interactions (*P*=.001), and "love" and "share" interactions (*P*=.004) can be explained more by the fan reach than by negative emotions. Social media contents with a high rate of "share" interaction may lead to low fan reach, which can be associated with a decrease in the rate of "love" and "comment" interactions. Overall, negative correlations between engagement indicators can indicate that the integrity of engagement should be assessed cautiously. The next research question was, "How does the strength of positive correlations change between engagement indicators?" Some highly significant correlations were observed between reactions, which indicated different Facebook users' combined interactions on the same content. The strongest highly significant positive correlations were found between "sad" and "angry" reactions (r_s=0.302, *P*<.001), "love" and "wow" reactions (r_s=0.204, *P*<.001), and "love" and "haha" reactions (r_s=0.141, *P*<.001). Other combinations of interactions on the given content may have arrived from different users or the same Facebook user. In this case, the strongest highly significant positive correlations were found between "comments" and "clicks" (r_s=0.417, *P*<.001), and between "comments" and reactions, especially "sad" (r_s=0.196, *P*<.001), "haha" (r_s=0.174, *P*<.001), and "angry" reactions (r_s=0.165, *P*<.001). Together, these results provide important insights into the type of combined reactions or interactions applied by Facebook users to react on smoking cessation support contents.

Second, the correlations related to negative Facebook interactions were analyzed. The last question in this research area was, "What is the relationship between engagement indicators and negative Facebook interactions?" We observed 3 correlations between some reactions ("like," "wow," and "sad") and negative Facebook interactions, which might be associated with negative emotional characteristics. A significant negative correlation was observed between the "like" reaction and negative interaction (*P*=.040), which can support its role among the engagement indicators. We found significant positive correlations between the "wow" reaction and negative interactions (*P*=.016), or between the "sad" reaction and negative interactions (*P*=.003), which can question their role among the engagement indicators. Another significant positive correlation was noted between "click" and negative interactions (*P*=.03), which can be explained more by a combination of interactions than by negative emotions. Facebook users might have collected more information by clicks before they applied negative interactions buttons. In summary, these results show that some engagement indicators correlated with negative Facebook interactions, which are activities against the smoking cessation intervention (eg, reports of spam or unlike of page).

Table 3. Spearman correlation matrix for the comparison between Facebook interactions regarding 1025 social media contents (N=1025).

	Reaction						Share	Comment	Click	Negative interactions
	Like	Love	Haha	Wow	Sad	Angry				
Reaction										
Like	1.00	0.008	0.063[a]	–0.077[a]	–0.120[b]	–0.136[b]	0.094[a]	–0.130[b]	0.147[b]	–0.064[a]
Love		1.00	0.141[b]	0.204[b]	0.063[a]	0.008	–0.090[a]	0.123[b]	–0.007	0.009
Haha			1.00	0.132[b]	0.140[b]	0.080[a]	–0.049	0.174[b]	0.002	0.052
Wow				1.00	0.104[a]	0.122[b]	–0.055	0.145[b]	0.087[a]	0.076[a]
Sad					1.00	0.302[b]	–0.011	0.196[b]	0.070[a]	0.091[a]
Angry						1.00	0.004	0.165[b]	0.097[a]	0.032
Share							1.00	–0.105[a]	0.097[a]	0.029
Comment								1.00	0.417[b]	0.059
Click									1.00	0.066[a]
Negative interactions										1.00

[a] Significant, *P*<.05 (2-tailed).

[b] Highly significant, *P*<.001 (2-tailed).

Discussion

Principal Results

The first aim in this research was to understand the association between organic reach and Facebook interactions during a smoking cessation intervention. Findings suggest that total organic reach may correlate positively with most Facebook interactions, except the "like" and "share" interactions. The highly significant negative correlation between total organic reach and the "like" reaction can highlight the widespread misconception of "more likes cause higher reach" [2]. The lack of a significant relationship between total organic reach and the "share" interaction may be explained by the unique function of "share." "Share" is the only interaction that has a direct effect on organic reach by sending the given content to others [24-26]. In this study, the "share" interaction correlated positively with nonfan reach, and negatively with fan reach. This positive correlation suggests that Facebook users who applied the "share" button usually sent the content to nonfan Facebook users, which led to higher nonfan reach. By contrast, the significant negative correlation between "shares" and fan reach highlights that the Facebook algorithm might decrease the fan reach in response to the notable increase in nonfan reach. Therefore, the "share" interaction may have a higher impact on content algorithmic ranking than the previous "page like," which is the basic requirement of fan reach. This mechanism can act as a "brake" of viral reach. Previous studies have showed that fan Facebook users are characterized by higher "share" activity than nonfans [2]. If "shares" would correlate positively with fan reach, a higher "share" activity of fan users could lead to a viral reach. This might be a possible explanation of the negative correlation between "shares" and nonfan reach as a "brake effect."

By contrast, the significant positive correlation between "clicks" and nonfan reach may be explained by an opposite cause–effect relationship. Nonfan Facebook users might apply "clicks" more often to gain new information than fan users. Nevertheless, the significant positive correlations between organic reach and other interactions can provide important insights into how to enhance total organic reach, especially fan reach. The "Haha" reaction, "comments" and the "love" reaction proved the most typical fan-specific interactions, which could also achieve the highest increase in total organic reach. Finally, an unanticipated finding was the significant positive correlation between organic reach and negative Facebook interactions, probably due to the fact that Facebook users could express their aversions related to the given content with negative interactions (eg, post hides, or reports of spam).

The second objective of the research was to assess the correlations between positive and negative emotional interactions during a Facebook-based smoking cessation intervention. Negative emotional reactions ("wow," "sad," "angry" reactions) and comments were sharply separated from the other engagement indicators. First, significant negative correlations were observed between the "like" reaction and "wow," "sad," "angry" reactions, and comments. Second, we found that a high rate of "wow" or "sad" reactions was associated with negative Facebook interactions. These results suggests that Facebook users could express their negative emotions with these interactions as a resistance to smoking cessation.

Finally, some special combinations of interactions are discussed. The correlations of the "share" interaction should be interpreted considering its special impact on fan reach. "Share" interactions may indirectly reduce fan reach and the rate of fan-specific interactions. This effect of "shares" can explain the significant negative correlations with "love" and "comment" interactions. In addition, the lack of a significant relationship with "haha," "wow," "sad," and "angry" reactions can be explained by the consequential low fan reach. The significant positive correlations of "shares" with "like" and "click" interactions might indicate frequently used combinations. Analyzing our data, the significant positive correlation between "click" and negative interactions was also a thought-provoking combined interaction, which may help to elucidate the reason why negative Facebook interactions could increase the organic reach.

Hypotheses for Future Research

This was an exploratory, hypothesis-generating study to get insights into the correlations of organic reach, engagement, and Facebook users' interactions. Textbox 3 presents some recommended hypotheses for more rigorous testing in the future. First, the hypotheses related to organic reach are discussed. We assume that the "like" reaction is likely to decrease the organic reach of smoking cessation support contents on a public Facebook page. Facebook users often used this interaction [2,11], perhaps this is why the Facebook algorithm may use it as a negative element in calculating organic reach. Future studies investigating the "like" reaction would be very interesting in Facebook-based smoking cessation interventions or other public health campaigns. It is also hypothesized that "share" interactions can increase the nonfan reach, and decrease the fan reach of smoking cessation support contents on a public Facebook page. This finding may have important implications for designing social media contents for a target population. In order to reach new (nonfan) Facebook users, "share" interactions on the given content should be encouraged. However, these social media contents probably fail to reach fan Facebook users according to our hypothesis. Further research to analyze the role of "shares" in fan reach and nonfan reach could provide more definitive evidence.

It is assumed that the fan reach would be increased by "haha" and "love" reactions and "comments," which would have the highest impact on the total organic reach of smoking cessation support contents on a public Facebook page. This finding may be a useful tool when creating Facebook posts to reach fan users during smoking cessation interventions or other public health campaigns. However, public health professionals who make Facebook posts should give suggestions for their audience about interactions carefully. Some techniques to increase engagement (called "engagement bait") should be avoided, because if Facebook detects this, it will automatically reduce the organic reach. In future investigations, it might be advantageous to use various contents to give suggestions with and without "engagement bait" for the Facebook page's audience about "like," "haha," "love" reactions, "shares," and "comments." It should be also noted that organic reach and interactions data for future studies are easily accessible in "Facebook Insights" for 6 months, or in the "Post Details" for more than 6 months. Further studies, which analyze these annual data of different public health Facebook pages, may help us to understand how the Facebook algorithm is modified each year, and public health interventions can follow these changes in the future. Facebook interactions are just one way of trying to improve organic reach of the given content in a public Facebook page. Future research should also consider other elements of the Facebook algorithm (eg, the timing of published content) and other parallel used deliveries (eg, Facebook groups, Facebook stories) [29].

Second, the hypotheses related to engagement indicators are discussed. An important issue for future research is to find which interactions can be classified as engagement indicators on Facebook [33]. Based on this study, it is hypothesized that negative Facebook interactions, negative emotional comments, and reactions ("wow," "sad," "angry") would reduce the engagement of Facebook-based smoking cessation interventions, because these interactions were separated from the other engagement indicators owing to the opposite direction of correlation. Furthermore, these results suggest that "wow" should not be considered a positive reaction, as it seems to have more similarities to a negative emotional response. It is also assumed that "like," "love," "haha" reactions, shares, positive comments, and clicks would raise the engagement of Facebook-based smoking cessation interventions. Future investigations should assess the impact of these Facebook interactions on other dimensions of engagement (eg, the amount or duration of usage) [29,33]. More broadly, research is also needed to determine the role of these interactions in other Facebook-based public health campaigns with different aims. For example, the "sad" reaction or negative emotional comment in addictology may indicate a resistance to the smoking cessation intervention, whereas the same interactions in cancer prevention may express Facebook users' engagement as a response to stories by patients with cancer [36].

Lastly, our findings suggest that some combined interactions (eg, "comments" and "clicks") would increase the probability of eliciting the other interaction (eg, a high number of "comments" can associate with the rise of "clicks"). This result is only relevant for Facebook-based smoking cessation interventions. In future investigations, self-report questionnaires might be used to explore the popular interactions of different public health target groups [29]. Despite its exploratory nature, this study can improve the practical implementation of these combined interactions. For example, if a public health professional wants to increase the rate of "clicks" (eg, to selecting a website), the usage of social media contents which generate "comments" may be more advantageous than the usage of contents which evoke reactions. This can be also an important issue for future research.

Textbox 3. Hypotheses for future testing based on this research.

Organic reach. Hypotheses for future testing are:

- "Like" reactions would decrease the organic reach of smoking cessation support contents on a public Facebook page.
- "Share" interactions would significantly increase the nonfan reach, and decrease the fan reach of smoking cessation support contents on a public Facebook page.
- "Haha," "love" reaction, and "comments" would have the highest impact on fan reach, and total organic reach of smoking cessation support contents on a public Facebook page.

Engagement indicators. Hypotheses for future testing are:

- Negative Facebook interactions, negative emotional comments, and reactions ("wow," "sad," "angry") would reduce the engagement of Facebook-based smoking cessation interventions.
- "Like," "love," "haha" reactions, shares, positive comments, and clicks would raise the engagement of Facebook-based smoking cessation interventions.
- Some combined interactions (eg, "comments" and "clicks") would increase the probability of eliciting the other interaction in a Facebook-based smoking cessation intervention.

Limitations

It should be noted that we used a convenience sample, and the audience of the investigated public Facebook page was heterogeneous. Furthermore, the availability of participants' demographic data is also limited to a post level because it could not be retrieved data from the "Post Details." That is why the results of the second correlational analysis cannot be generalized widely. Therefore, the correlations between the Facebook users' interactions can only be interpreted in the context of a smoking cessation intervention based on motivational interviewing.

However, these limitations might not affect the interpretation of the first correlational analysis because the relationship between the organic reach and Facebook interactions depends only on the "Facebook algorithm of content ranking." This algorithm is presumably free from demographic data or smoking status. Using age, gender, or other demographic data to reach users are against Facebook Community Standards. Furthermore, smoking status cannot be an element of the Facebook algorithm either because it is not registered in Facebook. Lastly, the Facebook algorithm probably evaluates the performance of the given post, rather than its public health content which supports smoking cessation or not. Therefore, the results of the first correlational analysis between the organic reach and Facebook interactions can be generalized for other Facebook-based interventions which use a public Facebook page. At the same time, these results should be interpreted with caution, regarding the other potential elements of the Facebook algorithmic content ranking (eg, mainly image-type post was included in this current research).

Finally, Facebook advertising can indirectly raise the organic reach and the number of interactions during the advertising period. The paid reach (66,463 people) was about double than organic reach (30,191 people) in the last month of our research period. However, this potential indirect growth in organic reach and the number of interactions affected all Facebook posts equally during the advertising period. Furthermore, boosted (paid) Facebook posts were excluded to avoid the direct effects of advertising.

Conclusions

The purpose of this research was to understand the relationship between Facebook users' interactions, organic reach, and engagement, which may be informative for further research and the design of Facebook-based smoking cessation interventions. The key strengths of this study are the large size of the data set and its long duration, which can shine new light on the pattern of change in the Facebook algorithm. Returning to the research questions described at the beginning of this paper, it is now possible to state that most Facebook interactions can correlate with the organic reach of a smoking cessation intervention positively, while the "like" reaction can correlate negatively. The strength of these correlations was proved to be different, which may mean variant emphases in the Facebook algorithm. This is the first study to report a disadvantage of the "like" reaction and highlight the advantages of other interactions (eg, the "haha" reaction or "comments") in algorithmic content ranking on Facebook. The analysis of fan reach and nonfan reach suggests the need for further categorization of fan-specific interactions (eg, "haha" or "love" reactions) and nonfan-specific interactions (eg, "shares" and "clicks"). The generalizability of these results is wide for Facebook-based public health interventions, because these correlations depend only the Facebook algorithm, which does not contain demographic data, smoking status, or other health risks. In addition, the Facebook algorithm of content ranking calculates with the performance of the given post, and it ignores the specific aim of the given public health intervention or the given theory or strategy of health behavioral change used in the Facebook post. This study made an attempt to explore the relationship between the Facebook algorithm and users' interactions; nevertheless, further research is needed to investigate other elements of the Facebook algorithm.

The correlational analysis of Facebook interactions raises some exciting hypotheses for future testing. A novel classification of engagement indicators should be considered in Facebook-based smoking cessation interventions. Negative Facebook interactions; negative emotional comments; and "wow," "sad," and "angry" reactions may decrease the engagement, whereas "like," "love," "haha" reactions, shares, positive comments, and clicks may increase the engagement of these interventions. Furthermore, other specific combinations of interactions can be useful to raise the probability of certain interactions under smoking cessation support contents. Based on our findings, we suggest implementing the continuous evaluation of Facebook interactions during interventions.

Acknowledgments

This research did not involve any costs or funding. All authors meet the conditions described in the recommendations by the International Committee of Medical Journal Editors Recommendations and gave final approval to the submitted paper. Ethics permission for this research was obtained from the Ethics Committee, Albert Szent-Györgyi Health Centre, University of Szeged. Reference number: 4794; 107/2020-SZTE.

Conflicts of Interest

None declared. The investigated Facebook page is a nonbusiness entity which employs volunteers and uses donations to sustain its mission. The data sets used and analyzed during this research are available from the corresponding author upon reasonable request.

Multimedia Appendix 1
Stimuli: examples of smoking cessation support contents based on motivational interviewing (doc).
[DOCX File , 19883 KB - jmir_v23i6e27853_app1.docx]

Multimedia Appendix 2
Exclusion criteria: examples of excluded social media contents (doc).
[DOCX File , 10667 KB - jmir_v23i6e27853_app2.docx]

Multimedia Appendix 3
Original data (xls).
[XLSX File (Microsoft Excel File), 219 KB - jmir_v23i6e27853_app3.xlsx]

References

1. Nabi RL, Prestin A, So J. Facebook friends with (health) benefits? Exploring social network site use and perceptions of social support, stress, and well-being. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):721-727. [doi: 10.1089/cyber.2012.0521] [Medline: 23790356]
2. Kite J, Foley BC, Grunseit AC, Freeman B. Please Like Me: Facebook and Public Health Communication. PLoS One 2016 Sep;11(9):e0162765 [FREE Full text] [doi: 10.1371/journal.pone.0162765] [Medline: 27632172]
3. Park H, Rodgers S, Stemmle J. Health Organizations' Use of Facebook for Health Advertising and Promotion. Journal of Interactive Advertising 2011 Sep;12(1):62-77. [doi: 10.1080/15252019.2011.10722191]
4. Kuipers M, Beard E, West R, Brown J. Associations between tobacco control mass media campaign expenditure and smoking prevalence and quitting in England: a time series analysis. Tob Control 2018 Jul;27(4):455-462 [FREE Full text] [doi: 10.1136/tobaccocontrol-2017-053662] [Medline: 28667091]
5. Durkin S, Brennan E, Wakefield M. Mass media campaigns to promote smoking cessation among adults: an integrative review. Tob Control 2012 Mar 16;21(2):127-138. [doi: 10.1136/tobaccocontrol-2011-050345] [Medline: 22345235]
6. Brown J, Kotz D, Michie S, Stapleton J, Walmsley M, West R. How effective and cost-effective was the national mass media smoking cessation campaign 'Stoptober'? Drug and Alcohol Dependence 2014 Feb;135:52-58. [doi: 10.1016/j.drugalcdep.2013.11.003]
7. Xu X, Alexander RL, Simpson SA, Goates S, Nonnemaker JM, Davis KC, et al. A cost-effectiveness analysis of the first federally funded antismoking campaign. Am J Prev Med 2015 Mar;48(3):318-325. [doi: 10.1016/j.amepre.2014.10.011] [Medline: 25498550]
8. Allom V, Jongenelis M, Slevin T, Keightley S, Phillips F, Beasley S, et al. Comparing the Cost-Effectiveness of Campaigns Delivered Various Combinations of Television and Online Media. Front Public Health 2018 Mar 23;6:83 [FREE Full text] [doi: 10.3389/fpubh.2018.00083] [Medline: 29629366]
9. Bold KW, Hanrahan TH, O'Malley SS, Fucito LM. Exploring the Utility of Web-Based Social Media Advertising to Recruit Adult Heavy-Drinking Smokers for Treatment. J Med Internet Res 2016;18(5):e107 [FREE Full text] [doi: 10.2196/jmir.5360] [Medline: 27194456]
10. Thackeray R, Neiger BL, Smith AK, Van Wagenen SB. Adoption and use of social media among public health departments. BMC Public Health 2012;12:242 [FREE Full text] [doi: 10.1186/1471-2458-12-242] [Medline: 22449137]

11. Kite J, Grunseit A, Li V, Vineburg J, Berton N, Bauman A, et al. Generating Engagement on the Make Healthy Normal Campaign Facebook Page: Analysis of Facebook Analytics. JMIR Public Health Surveill 2019 Jan 14;5(1):e11132 [FREE Full text] [doi: 10.2196/11132] [Medline: 31344679]
12. Andrade EL, Evans WD, Barrett N, Edberg MC, Cleary SD. Strategies to Increase Latino Immigrant Youth Engagement in Health Promotion Using Social Media: Mixed-Methods Study. JMIR Public Health Surveill 2018 Dec 19;4(4):e71 [FREE Full text] [doi: 10.2196/publichealth.9332] [Medline: 30567689]
13. Ford K, Albritton T, Dunn T, Crawford K, Neuwirth J, Bull S. Youth Study Recruitment Using Paid Advertising on Instagram, Snapchat, and Facebook: Cross-Sectional Survey Study. JMIR Public Health Surveill 2019 Oct 09;5(4):e14080 [FREE Full text] [doi: 10.2196/14080] [Medline: 31599739]
14. Barklamb A, Molenaar A, Brennan L, Evans S, Choong J, Herron E, et al. Learning the Language of Social Media: A Comparison of Engagement Metrics and Social Media Strategies Used by Food and Nutrition-Related Social Media Accounts. Nutrients 2020 Sep 16;12(9):A [FREE Full text] [doi: 10.3390/nu12092839] [Medline: 32948033]
15. Chan L, O'Hara B, Phongsavan P, Bauman A, Freeman B. Review of Evaluation Metrics Used in Digital and Traditional Tobacco Control Campaigns. J Med Internet Res 2020 Aug 11;22(8):e17432 [FREE Full text] [doi: 10.2196/17432] [Medline: 32348272]
16. Theiss SK, Burke RM, Cory JL, Fairley TL. Getting Beyond Impressions: An Evaluation of Engagement with Breast Cancer-related Facebook Content. Mhealth 2016 Nov 07;2:41-41 [FREE Full text] [doi: 10.21037/mhealth.2016.10.02] [Medline: 27840816]
17. Neiger BL, Thackeray R, Van Wagenen SA, Hanson CL, West JH, Barnes MD, et al. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. Health Promot Pract 2012 Mar;13(2):159-164. [doi: 10.1177/1524839911433467] [Medline: 22382491]
18. Hefler M, Kerrigan V, Grunseit A, Freeman B, Kite J, Thomas D. Facebook-Based Social Marketing to Reduce Smoking in Australia's First Nations Communities: An Analysis of Reach, Shares, and Likes. J Med Internet Res 2020 Dec 10;22(12):e16927 [FREE Full text] [doi: 10.2196/16927] [Medline: 33300883]
19. Campbell I, Rudan I. Analysis of public engagement with ten major global health topics on a social network profile and a newspaper website. J Glob Health 2020 Jun;10(1):010902 [FREE Full text] [doi: 10.7189/jogh.10.010902] [Medline: 32257171]
20. Merchant G, Weibel N, Patrick K, Fowler JH, Norman GJ, Gupta A, et al. Click "like" to change your behavior: a mixed methods study of college students' exposure to and engagement with Facebook content designed for weight loss. J Med Internet Res 2014 Jun 24;16(6):e158 [FREE Full text] [doi: 10.2196/jmir.3267] [Medline: 24964294]
21. Rus HM, Cameron LD. Health Communication in Social Media: Message Features Predicting User Engagement on Diabetes-Related Facebook Pages. Ann Behav Med 2016 Dec;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
22. Lee Y, Phua J, Wu T. Marketing a health Brand on Facebook: Effects of reaction icons and user comments on brand attitude, trust, purchase intention, and eWOM intention. Health Mark Q 2020 Apr 14;37(2):138-154. [doi: 10.1080/07359683.2020.1754049] [Medline: 32286189]
23. Kim SJ, Marsch LA, Brunette MF, Dallery J. Harnessing Facebook for Smoking Reduction and Cessation Interventions: Facebook User Engagement and Social Support Predict Smoking Reduction. J Med Internet Res 2017 May 23;19(5):e168 [FREE Full text] [doi: 10.2196/jmir.6681] [Medline: 28536096]
24. Zhang N, Tsark J, Campo S, Teti M. Facebook for Health Promotion: Female College Students' Perspectives on Sharing HPV Vaccine Information Through Facebook. Hawaii J Med Public Health 2015 Apr;74(4):136-140 [FREE Full text] [Medline: 25954600]
25. Obamiro K, West S, Lee S. Like, comment, tag, share: Facebook interactions in health research. Int J Med Inform 2020 May;137:104097. [doi: 10.1016/j.ijmedinf.2020.104097] [Medline: 32092667]
26. Buttle F, Groeger L. Who says what to whom in what channel? A rules theoretic perspective on word-of-mouth marketing. Journal of Marketing Management 2017 May 19;33(13-14):1035-1059. [doi: 10.1080/0267257X.2017.1325390]
27. Struik LL, Baskerville NB. The role of Facebook in Crush the Crave, a mobile- and social media-based smoking cessation intervention: qualitative framework analysis of posts. J Med Internet Res 2014 Jul;16(7):e170 [FREE Full text] [doi: 10.2196/jmir.3189] [Medline: 25016998]
28. Bennetts SK, Hokke S, Crawford S, Hackworth NJ, Leach LS, Nguyen C, et al. Using Paid and Free Facebook Methods to Recruit Australian Parents to an Online Survey: An Evaluation. J Med Internet Res 2019 Mar 06;21(3):e11206 [FREE Full text] [doi: 10.2196/11206] [Medline: 30839282]
29. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun 01:257-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
30. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. J Med Internet Res 2006 Aug;8(3):e15 [FREE Full text] [doi: 10.2196/jmir.8.3.e15] [Medline: 16954125]

31. Couper MP, Alexander GL, Zhang N, Little RJA, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
32. Cole-Lewis H, Perotte A, Galica K, Dreyer L, Griffith C, Schwarz M, et al. Social Network Behavior and Engagement Within a Smoking Cessation Facebook Page. J Med Internet Res 2016 Aug 02;18(8):e205 [FREE Full text] [doi: 10.2196/jmir.5574] [Medline: 27485315]
33. Platt T, Platt J, Thiel DB, Kardia SLR. Facebook Advertising Across an Engagement Spectrum: A Case Example for Public Health Communication. JMIR Public Health Surveill 2016 May 30;2(1):e27 [FREE Full text] [doi: 10.2196/publichealth.5623] [Medline: 27244774]
34. CigiSzünet. URL: www.facebook.com/cigiszunet [accessed 2021-06-14]
35. Pócs D, Kovács R, Óvári T, Erdős C, Kelemen O. [Tobacco reduction on Facebook among 14-35-year-olds]. Orv Hetil 2019 Feb;160(6):220-227. [doi: 10.1556/650.2019.31314] [Medline: 30730203]
36. Tang C, Zhou L, Plasek J, Rozenblum R, Bates D. Comment Topic Evolution on a Cancer Institution’s Facebook Page. Appl Clin Inform 2017 Dec 20;08(03):854-865. [doi: 10.4338/aci-2017-04-ra-0055]

Edited by G Eysenbach; submitted 09.02.21; peer-reviewed by W Evans, A Grunseit, L Chan, S Tomczyk; comments to author 03.03.21; revised version received 09.04.21; accepted 29.04.21; published 21.06.21.

Please cite as:
Pócs D, Adamovits O, Watti J, Kovács R, Kelemen O
Facebook Users’ Interactions, Organic Reach, and Engagement in a Smoking Cessation Intervention: Content Analysis
J Med Internet Res 2021;23(6):e27853
URL: https://www.jmir.org/2021/6/e27853
doi:10.2196/27853
PMID:34152280

Original Paper

Machine Learning Analysis to Identify Digital Behavioral Phenotypes for Engagement and Health Outcome Efficacy of an mHealth Intervention for Obesity: Randomized Controlled Trial

Meelim Kim[1*], MA; Jaeyeong Yang[2*], MS; Woo-Young Ahn[2,3], PhD; Hyung Jin Choi[1,4], MD, PhD

[1]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea
[2]Department of Psychology, Seoul National University, Seoul, Republic of Korea
[3]Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Republic of Korea
[4]Department of Anatomy and Cell Biology, Neuroscience Research Institute, Wide River Institute of Immunology, Gangwon-do, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Hyung Jin Choi, MD, PhD
Department of Biomedical Sciences
Seoul National University College of Medicine
28 Yungun-Dong, Chongno-Gu
Seoul
Republic of Korea
Phone: 82 27408204
Email: hjchoi@snu.ac.kr

Abstract

Background: The digital health care community has been urged to enhance engagement and clinical outcomes by analyzing multidimensional digital phenotypes.

Objective: This study aims to use a machine learning approach to investigate the performance of multivariate phenotypes in predicting the engagement rate and health outcomes of digital cognitive behavioral therapy.

Methods: We leveraged both conventional phenotypes assessed by validated psychological questionnaires and multidimensional digital phenotypes within time-series data from a mobile app of 45 participants undergoing digital cognitive behavioral therapy for 8 weeks. We conducted a machine learning analysis to discriminate the important characteristics.

Results: A higher engagement rate was associated with higher weight loss at 8 weeks ($r=-0.59$; $P<.001$) and 24 weeks ($r=-0.52$; $P=.001$). Applying the machine learning approach, lower self-esteem on the conventional phenotype and higher in-app motivational measures on digital phenotypes commonly accounted for both engagement and health outcomes. In addition, 16 types of digital phenotypes (ie, lower intake of high-calorie food and evening snacks and higher interaction frequency with mentors) predicted engagement rates (mean R^2 0.416, SD 0.006). The prediction of short-term weight change (mean R^2 0.382, SD 0.015) was associated with 13 different digital phenotypes (ie, lower intake of high-calorie food and carbohydrate and higher intake of low-calorie food). Finally, 8 measures of digital phenotypes (ie, lower intake of carbohydrate and evening snacks and higher motivation) were associated with a long-term weight change (mean R^2 0.590, SD 0.011).

Conclusions: Our findings successfully demonstrated how multiple psychological constructs, such as emotional, cognitive, behavioral, and motivational phenotypes, elucidate the mechanisms and clinical efficacy of a digital intervention using the machine learning method. Accordingly, our study designed an interpretable digital phenotype model, including multiple aspects of motivation before and during the intervention, predicting both engagement and clinical efficacy. This line of research may shed light on the development of advanced prevention and personalized digital therapeutics.

Trial Registration: ClinicalTrials.gov NCT03465306; https://clinicaltrials.gov/ct2/show/NCT03465306

(J Med Internet Res 2021;23(6):e27218) doi:10.2196/27218

KEYWORDS

digital phenotype; clinical efficacy; in-app engagement; machine learning analysis; mobile phone

Introduction

Background

The use of mobile tools, such as smartphones, to assist health care systems is rapidly growing in the current era. As the interactions between individuals and digital communities via mobile devices are progressively embedded in human lives, understanding the concept of a *digital phenotype* is also important. A digital phenotype is a collected set of data in a digital system intentionally demonstrated by humans or as a secondary outcome of other activities, influencing human behavior. Specifically, the expanding body of health-related data from mobile devices allows us to address real-world life events with problematic behaviors. For example, data related to the timing and periods of one's digital footprint can be examined as part of a patient's features with insomnia [1]. Similarly, data from Google searches can recognize suicidal ideation [2]. To date, digital technologies such as smartphone apps afford moment-by-moment perceptible measurements of a person's behavior regarding preventive and predictive ways to manage health.

Obtaining app users' attention is a critical issue related to the app's potential efficacy for behavior change. The association between intervention exposure and efficacy emphasizes the need for a detailed understanding of user engagement [3]. When we deliver an intervention via a mobile app, the users must actively and frequently engage with mobile apps to succeed within the treatment. Thus, identifying predictive markers that can inform engagement in mobile health (mHealth) interventions could potentially strengthen its effectiveness. Previous studies have found that the involvement of social and gamified components or offering personalized feedback from human factors effectively enhances user engagement for app-based interventions [4,5]. In fact, identifying the major principles that can predict users' engagement and health outcomes is important for exploring systemic elements to strengthen user engagement in digital intervention. Engagement with digital technology is intricate because it is not stationary but a progressive process [6]. It is also multifaceted in its environment, reflecting the quality of the user's practice, their communication features, and their willingness to use the app over time or repeatedly [7]. Of special interest to this issue, it is noted that intrinsic motivation is a significant precursor for engagement [8]. Moreover, a wide range of cognitive and emotional states, such as self-interest and self-efficacy, are closely related to the user's engagement [7]. Therefore, it is important to examine motivation, behavior, emotion, and cognition to understand the changes in users' engagement and predict clinical outcomes. This will intensify the treatment's efficacy and find good responders to precision medicine. However, finding the major indicator that predicts who will benefit the most from a digital intervention is insufficient. This resulted in only a minor portion of users obtaining advantages from the digital health care system [9,10]. Thus, it is necessary to explore how comprehensive and multidimensional digital phenotypes detect individual differences and determine user engagement in digital interventions.

Another major issue in the digital era is the interpretation and filtering of data for clinical decisions. Although the rapid growth of digital technologies has led to comprehensive and abundant information about one's health status, analytical methods to clarify and simplify it have not advanced at a compatible pace [11]. This could be addressed as the main bottleneck in current digital phenotyping studies. Some pioneering research has demonstrated statistical methods to derive insights (which predict outcomes) from various digital phenotypes [12-14]. However, the data are mostly heterogeneous and mixed with structured and unstructured frames containing random sampling, artifacts, and inconsistent completion, making traditional statistical models difficult. This can lead to limited or biased results from the data and a lack of replicability of the conclusions. Compared with conventional analytical methods, machine learning analysis can obtain information from scattered and intricate data, offering insights to promote clinical decision making. A recent study has shown that mortality prediction models using intensive care unit data based on a machine learning approach are superior to conventional methods [15]. Algorithms supporting individual-specific predictions may enhance the usability of machine learning prediction models. This could aid in the adaptation of machine learning models as clinical decision-support tools.

Objectives

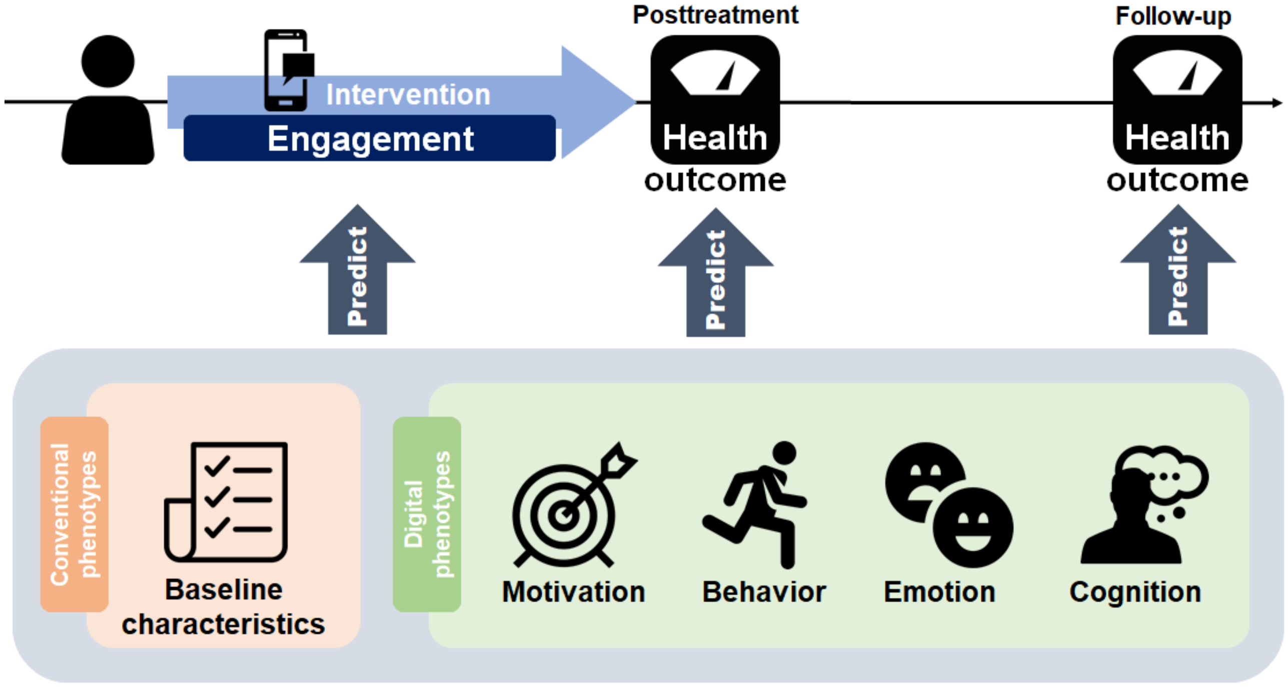
In this study, we aim to investigate multidimensional information at different time points using various assessment methods to monitor and predict the engagement and efficacy of the primary outcome. This study plays a significant role in establishing the most practical and effective mHealth intervention paradigm.

Methods

Study Design and Participants

We performed a post hoc analysis based on data from a previously reported open-label, 8-week, active comparator randomized controlled trial in the digital cognitive behavioral therapy (dCBT) study. The trial was registered with ClinicalTrials.gov (NCT03465306) in March 2018. Methods of recruitment, inclusion and exclusion criteria, and demographics have been published elsewhere [16]. All study participants provided written informed consent before enrollment in the study. The Institutional Review Board of Seoul National University Hospital approved this study (H-1707-122-872). The study protocol was registered at ClinicalTrials.gov (NCT03465306) on January 15, 2018. This study was conducted to validate the clinical efficacy of the obesity dCBT model and to identify factors related to its efficacy. Furthermore, all the digital phenotypes were averaged for each participant to predict their engagement during the intervention and their health outcomes for both the short term (8 weeks) and long term (24 weeks). A conceptual framework of mHealth components, including examples of digital phenotypes, is presented in Figure 1.

Figure 1. A conceptual framework of mobile health components and examples of digital phenotypes.



A total of 70 female participants aged between 18 and 39 years, with a BMI of 25-40, eligibility for smartphone use (assessed during the screening interview), and scores in the highest 40% on the Situational Motivation Scale (SIMS; scores above 68 out of 112 in total) were enrolled. We analyzed only 45 participants from the dCBT group. No analysis was performed in the control group. Among the dCBT group, we excluded 6 participants due to dropout, 1 participant due to withdrawal, and 1 participant due to lack of participation (less than 15 days). Therefore, the data analyzed included 37 participants.

Randomization and Masking

Participants were randomly assigned to a control group or dCBT group with a ratio of 1 to 2 to amplify the power of the dCBT group within analysis and to administer a more robust test within resource restraint. Research participants and research staff were aware of group assignments, but the group assignments were blinded to the technicians and clinical staff.

Procedures

The aim of this analysis was a post hoc analysis of the effects of a dCBT intervention on obesity. The detailed design and procedures have been described elsewhere [16]. Briefly, the participants in the dCBT group (app+human cognitive behavioral therapy) consisted of daily individualized feedback and assignments from a clinical psychologist based on the CBT modules for 8 weeks. CBT contents were obtained from the program used by the clinicians' guidelines [17]. The therapist monitored multidimensional components related to the behavior, cognition, emotion, and motivation of each participant in the dCBT group. In contrast, the participants in the control group (app only) were instructed to use the food diary by themselves. All participants were asked to visit at baseline and at 8 and 24 weeks. Anthropometric and self-administered questionnaires were collected at each study visit. The Noom app was mainly used to log food diaries and deliver messages between the therapist and participants.

Measures

The statistical information for the baseline characteristics and in-app measures is presented in Table 1. There are two main structures: conventional and digital phenotypes, which are classified based on different algorithms. Conventional phenotypes were composed of previously developed and validated surveys. Digital phenotypes are generated by a newly devised scoring system consisting of a combination of active and passive digital features gathered from digital devices. These phenotypes are categorized into four different dimensions: behavioral, cognitive, emotional, and motivational. For conventional phenotypes, four indices for each behavioral and emotional dimension and one index for cognitive and motivational dimensions were assessed. A total of 17 indices for behavioral, 1 for cognitive, 5 for emotional, and 4 for motivational dimensions were assessed regarding digital phenotypes. These categorizations among the four dimensions were proposed based on previous studies [18-21]. The surveys for each dimension were also developed and not validated because they were used only to monitor the users' condition and not for clinical diagnosis.

Table 1. Participant characteristics on demographic, behavioral, cognitive, emotional, and motivational measures.

Phenotype	Value, mean (SD)
Demographic information	
Age (years)	22.59 (3.68)
Presession BMI	27.86 (3.14)
Postsession BMI	27.01 (3.51)
Conventional phenotypes	
Behavioral	
Restricted eating (DEBQ-RE[a])	29.81 (6.90)
Emotional eating (DEBQ-EM[b])	37.54 (9.85)
Environmental eating (DEBQ-ENV[c])	34.76 (4.82)
Food addiction (YFAS[d])	2.54 (1.30)
Cognitive	
Automatic thoughts (ATQ-30[e])	56.92 (21.81)
Emotional	
Depression (BDI[f])	13.22 (8.04)
Anxiety (TAI[g])	47.92 (10.03)
Body satisfaction (BSQ-8C[h])	35.84 (7.30)
Self-esteem (RSES[i])	19.65 (5.15)
Motivational	
Conventional motivation (SIMS[j])	75.97 (5.89)
Digital phenotypes	
Behavioral	
Carbohydrate	142.95 (26.49)
Protein	49.69 (10.60)
Fat	38.46 (9.37)
Sodium	2190.52 (585.95)
Sugar	39.47 (11.42)
Breakfast	201.03 (108.33)
Morning snack	18.28 (15.98)
Lunch	402.16 (98.56)
Afternoon snack	56.71 (39.97)
Dinner	438.98 (120.26)
Evening snack	67.98 (56.61)
High-calorie food	0.29 (0.09)
Moderate calorie food	0.48 (0.06)
Low-calorie food	0.18 (0.09)
Steps	6485.00 (2618.54)
Exercise	8.17 (8.14)
Interaction frequency	9.48 (2.34)
Cognitive	
Obesity automatic thoughts	0.49 (0.64)

Phenotype	Value, mean (SD)
Emotional	
Irritated	46.53 (23.39)
Lonely	49.43 (24.52)
Nervous	47.26 (23.81)
Bored	47.74 (24.18)
Depressed	47.04 (24.27)
Motivational	
Will	4.58 (2.23)
Importance	3.73 (2.10)
Confidence	4.11 (2.20)
Satisfaction	4.46 (2.40)

[a] DEBQ-RE: Dutch Eating Behavior Questionnaire–Restricted Eating.

[b] DEBQ-EM: Dutch Eating Behavior Questionnaire–Emotional Eating.

[c] DEBQ-ENV: Dutch Eating Behavior Questionnaire–Environmental Eating.

[d] YFAS: Yale Food Addiction Scale.

[e] ATQ-30: Automatic Thoughts Questionnaire-30.

[f] BDI: Beck Depression Inventory.

[g] TAI: Trait Anxiety Inventory.

[h] BSQ-8C: Body Shape Questionnaire-8C.

[i] RSES: Rosenberg Self-Esteem Scale.

[j] SIMS: Situational Motivational Scale.

Participants' situational motivation toward the weight loss program was assessed using an adapted version of the SIMS. SIMS typically measures four types of motivation to engage in a task (herein, the weight loss program) at a specific point in time, with four items per subscale: intrinsic motivation, identified regulation, external regulation, and motivation. SIMS has demonstrated acceptable levels of reliability and validity in previous studies. The Body Shape Questionnaire-8C (BSQ-8C) is a brief form of the BSQ and consists of eight items extracted from the full version measuring the extent of the psychopathology of concerns about body shape. Higher BSQ values indicated greater body dissatisfaction. Depression was assessed using the Korean version of the Beck Depression Inventory scoring system. A total score from 0 to 9 indicated no depression, 10 to 15 indicated mild depression, 16 to 23 indicated moderate depression, and 24 to 63 indicated severe depression. Anxiety was measured using the 20-item Trait Anxiety Scale of the State-Trait Anxiety Inventory, with greater scores indicating more trait anxiety. The Rosenberg Self-Esteem Scale measure of self-esteem was used in this research with a 10-item scale with all negatively worded items. Thus, higher scores implied lower self-esteem. Eating behavior was measured using the Dutch Eating Behavior Questionnaire, which has three different psychologically based eating behaviors: restrained eating, emotional eating, and external eating. It contains 33 items, with higher scores indicating a greater tendency to present subscale behavior. The frequency of automatic negative thoughts associated with depression was assessed using the Automatic Thoughts Questionnaire-30. The scores ranged from 30 to 150, and higher scores implied that the participants experienced automatic negative thoughts more often. All psychological questionnaires were presented in Korean.

Six types of behavioral phenotypes, modified and extended from previous studies [22,23], were assessed in apps: food restriction, overeating and binge eating, late-night meals, snacking, food choice, and activity rate. Food restriction was evaluated using calories per meal per day. Overeating and binge eating were assessed by calories per meal per day and the speed per meal—the late-night meal was investigated using the dinner calories and the time per meal. Snacking was estimated using snack calories. Food choice was examined based on the type of food per meal, total amount of sodium and sugar, number of food types per meal, and percentage of nutritional types (carbohydrate, protein, and fat). The activity rate was measured as the number of steps and the total hours of exercise. Automatic thoughts were grouped into six categories: selective abstraction, arbitrary inference, overgeneralization, magnification or minimization, personalization, and absolutism. There were 20 automatic thoughts, and participants could add thoughts related to food or eating behaviors. Example statements for automatic thoughts are listed in Table S1 in Multimedia Appendix 1. We assessed 5 negative emotions closely related to problematic eating habits: irritation, loneliness, nervousness, boredom, and depression. The participants were asked to report each type of negative emotion score using a visual analog scale between 0 and 100. Motivation was assessed using four dimensions: will, rank of importance, confidence, and satisfaction. These different types of motivation were scored using a 10-point Likert scale (1-10).

Outcomes

The primary outcomes were changes in body weight and number logged into the app. Body weight was assessed by InBody H20B (InBody Co, Ltd) at baseline and 8 and 24 weeks in light street clothing and without socks and shoes. The number logged into the app was examined by tracking the actions such as responses to the daily assessment (responses per day), meals logged (meals per week), green foods defined by Noom (logged per week), exercise logged (times per week), exercise time registered (minutes per week), steps recorded (steps per week), weigh-ins logged (times per week), articles read (articles per week), group posts (posts per week), group comments (comments per week), messages sent to coaches (messages per week), and group likes (likes per week). The engagement rate was assessed using these objective indices for each participant.

Statistical Analysis

We analyzed the data to predict three target outcomes: (1) the number of mobile activities during the experiment session, (2) the weight change rate between presession (week 0) and postsession (week 8), and (3) the weight change rate between postsession and follow-up. The weight change rates were calculated as the ratio of the weight difference to the baseline weight as $(\text{weight}_{\text{before}}–\text{weight}_{\text{after}})/\text{weight}_{\text{before}}$. Correlations between the number of logs and weight change rates were analyzed to determine the relationship between engagement and health outcomes.

A machine learning approach using an elastic net was conducted [24]. The elastic net is a penalized regression method that automatically selects significant variables by reducing the regression coefficients of unimportant features to zero. Using 37 behavioral, cognitive, motivational, and emotional measures, we tried to reveal which measure contributes to predicting behavioral changes before and after treatment.

The analysis procedure for the out-of-sample regressions was similar to that in a previous study [25,26]. To conduct out-of-sample regression, we used leave-one-out cross-validation, which trains a model with data except for a single point and then evaluates the point's prediction. The root mean squared errors (RMSE) computed for all possible train test splits are averaged to the leave-one-out cross-validation error, which is a measure for evaluating the model fit.

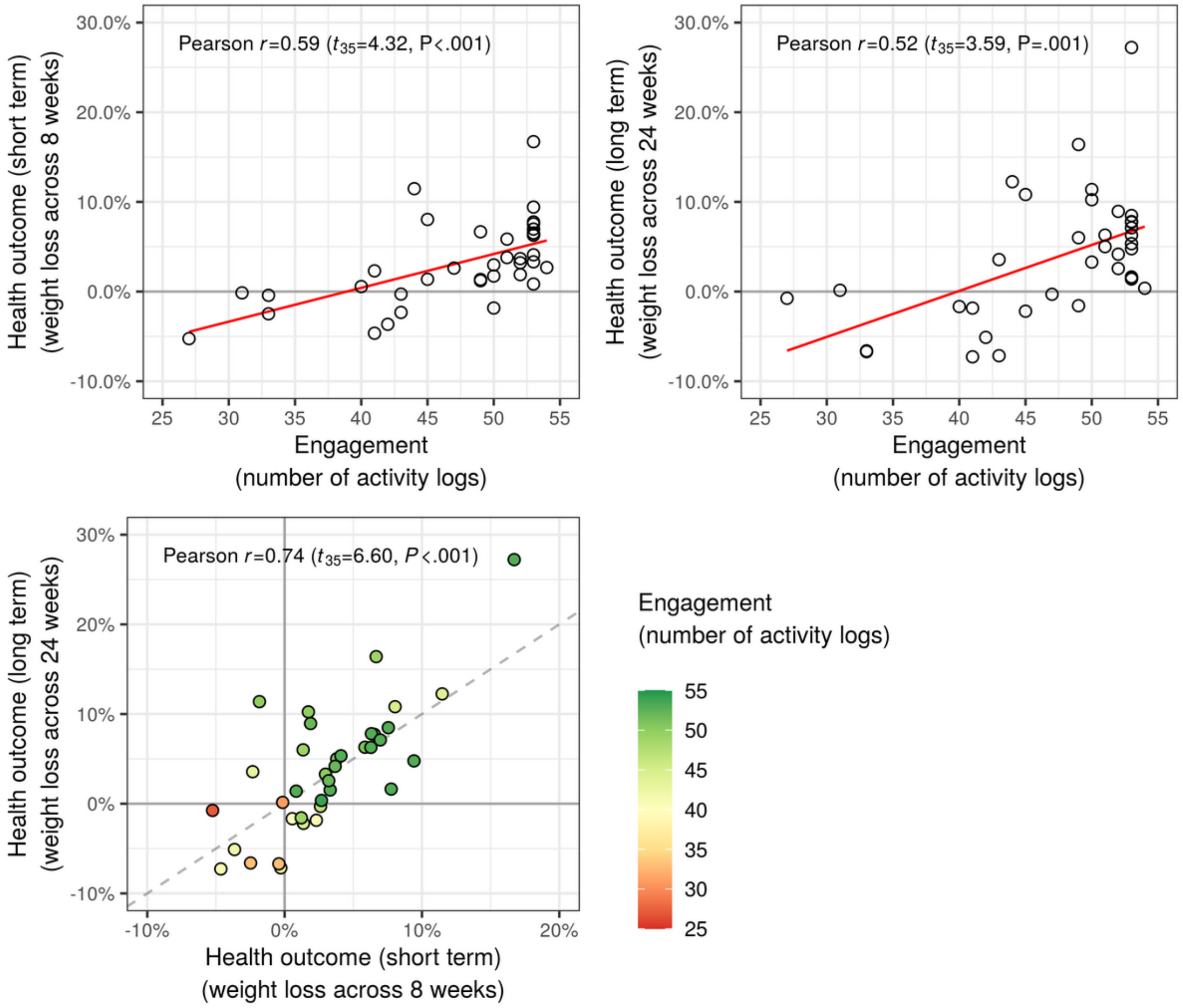
To acquire generalizable coefficients, we conducted model fitting 1000 times for each possible α value, which is the ratio between the ridge and lasso penalty terms. The number of iterations was chosen according to previous literature using a similar approach [25,26]. Figure S1 in Multimedia Appendix 1 shows the RMSE with 100 α values (from 0.01 to 1 with an interval of 0.01), and we chose the α value that minimizes RMSE across all participants. Then, to identify predictors for engagement and health outcomes, we computed mean β coefficients across 1000 iterations, and only phenotypes that were significant in more than 5% of 1000 iterations were selected as predictors for each model [25,26].

Results

Relationship Between the Number of Logs and Weight Changes

Figure 2 shows the correlations between the number of logs (engagement) and weight change (health outcomes). For the weight change during the 8-week intervention, two variables were highly correlated (r=−0.59; two-tailed t_{35}=−4.32; P<.001; Figure 2), which indicates that participants who had engaged in the in-app activity more actively lost weight. This result was the same for the weight change between baseline and follow-up (r=−0.52; two-tailed t_{35}=−3.59; P<.001). These short-term and long-term health outcomes were highly correlated (r=0.74; two-tailed t_{35}=6.60; P<.001).

Figure 2. Relationships between engagement and health outcomes. The health outcome larger than zero indicates weight loss compared with baseline.



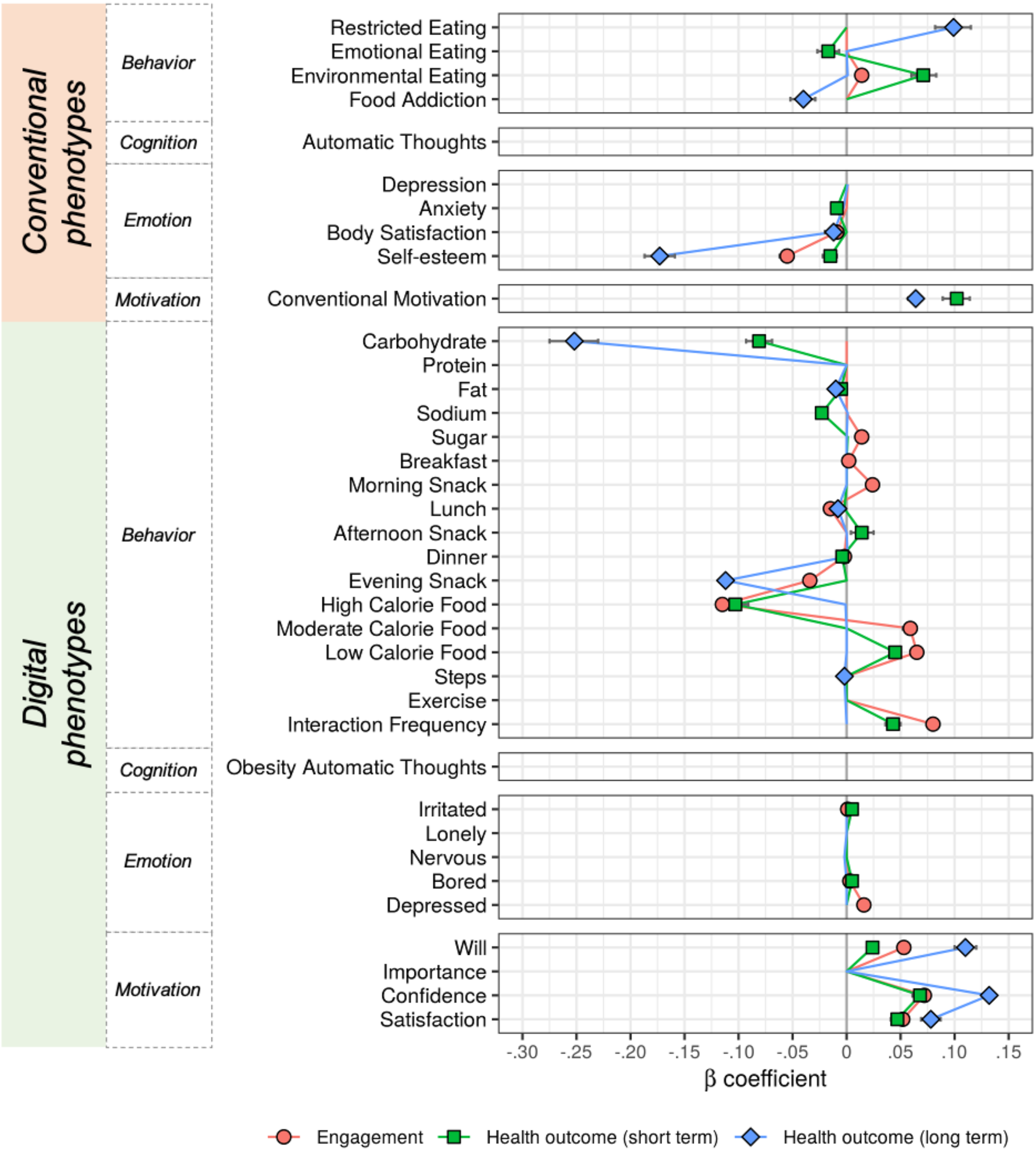
In addition, we tried to check whether there exist nonlinear relationships between the log number of activity logs and short-and long-term health outcomes. Such log transformations did not show significant differences for the relationship with short-term health outcomes (r_{log}=0.58; z_{34}=0.7843; *P*=.43) and even for those with long-term health outcomes (r_{log}=0.50; z_{34}=1.4579; *P*=.14).

Elastic Net Results

Through the leave-one-out cross-validations with different values for the mixing parameter (α), we chose the best value for each model that showed the minimum RMSE between the data and predicted outcomes. The estimated mixing parameters, α, were .08, .15, and .53 for predicting engagement, short-term health outcome, and long-term health outcome, respectively (Figure S1 in Multimedia Appendix 1). The α estimate for the long-term health outcome was much higher than that in the other two models, suggesting that the multivariate pattern is more parsimonious. Its coefficients are prone to shrink to zero while predicting long-term weight changes.

Figure 3 illustrates the multivariate profiles of conventional and digital phenotypes to predict in-app engagement and the health outcomes of digital health care. In-app engagement, computed as the number of daily activity logs, was significantly associated with lower self-esteem, lower body satisfaction, and higher external eating behaviors, measured as conventional phenotypes. For digital phenotypes, engagement was predicted by lower intake of food with a high calorie density index (CDI), higher food intake in the morning (breakfast and morning snack), lower food intake after that (lunch, dinner, and evening snack), higher sugar intake, higher intake of moderate or low CDI food, and higher frequency of interactions with the therapist. Higher emotional and motivational measures in digital phenotypes were also involved, such as irritation, boredom, depression, satisfaction, will, and confidence.

Figure 3. Multivariate patterns of conventional and digital phenotypes for predicting engagement (red) as well as short-term (green) and long-term (blue) health outcomes. Points indicate the averaged β coefficients across 100 repetitions of net elastic analysis (see the Methods section for details). A positive β estimate of a phenotype indicates an association between the phenotype and higher in-app activities (engagement) or more weight loss (health outcomes). The points, which contain zero in the simulated 95% ranges, are omitted.



For short-term health outcomes, lower emotional eating behavior, lower self-esteem, lower anxiety, higher external eating behavior, and higher motivation predicted the weight change rate for 8 weeks. The 8-week weight change was also predicted by lower intake of high CDI food, lower carbohydrate, lower sodium, lower fat intake, higher afternoon snack intake, lower dinner intake, higher intake of low CDI food, and higher frequency interactions with a health care mentor. Furthermore, short-term health outcomes were positively associated with emotional and motivational features in digital phenotypes, such as boredom, irritation, will, satisfaction, and confidence.

In contrast, fewer phenotypes are involved in the prediction of long-term health outcomes. Lower self-esteem, lower food addiction, lower body satisfaction, higher motivation, and higher restricting eating behavior in conventional phenotypes predicted the 24-week weight change. For digital phenotypes, the long-term health outcome was predicted by lower carbohydrate intake, lower lunch and evening snack intake, lower fat intake, lower steps in a day, higher satisfaction, higher will, and higher confidence.

Common predictors across dependent variables were associated with different phenotypes (Figure 4 and Table 2). Engagement and health outcomes were commonly affected by lower self-esteem in conventional phenotypes and higher in-app motivational measures in digital phenotypes. In other words, decreased self-esteem before the intervention and inclined motivation during the intervention highly predicted more in-app activities and more weight loss following the intervention. Furthermore, common predictors between engagement and short-term health outcomes include the behavioral dimension of digital phenotypes, such as the frequency of coach interaction and low- or high-calorie food intake. Carbohydrate intake was the most commonly influential predictor of short-term and long-term health outcomes. Conversely, conventional and digital phenotypes' motivational measures were positively associated with health outcomes (Figure 5).

Figure 4. Common predictors between engagement and health outcomes for (A) health outcome (short term) versus engagement, (B) health outcome (long term) versus engagement, and (C) health outcome (long term) versus health outcome (short term). Each axis indicates the β estimate for predicting engagement and health outcomes. A positive β coefficient indicates a positive association with engagement but negative associations with health outcomes (weight changes).

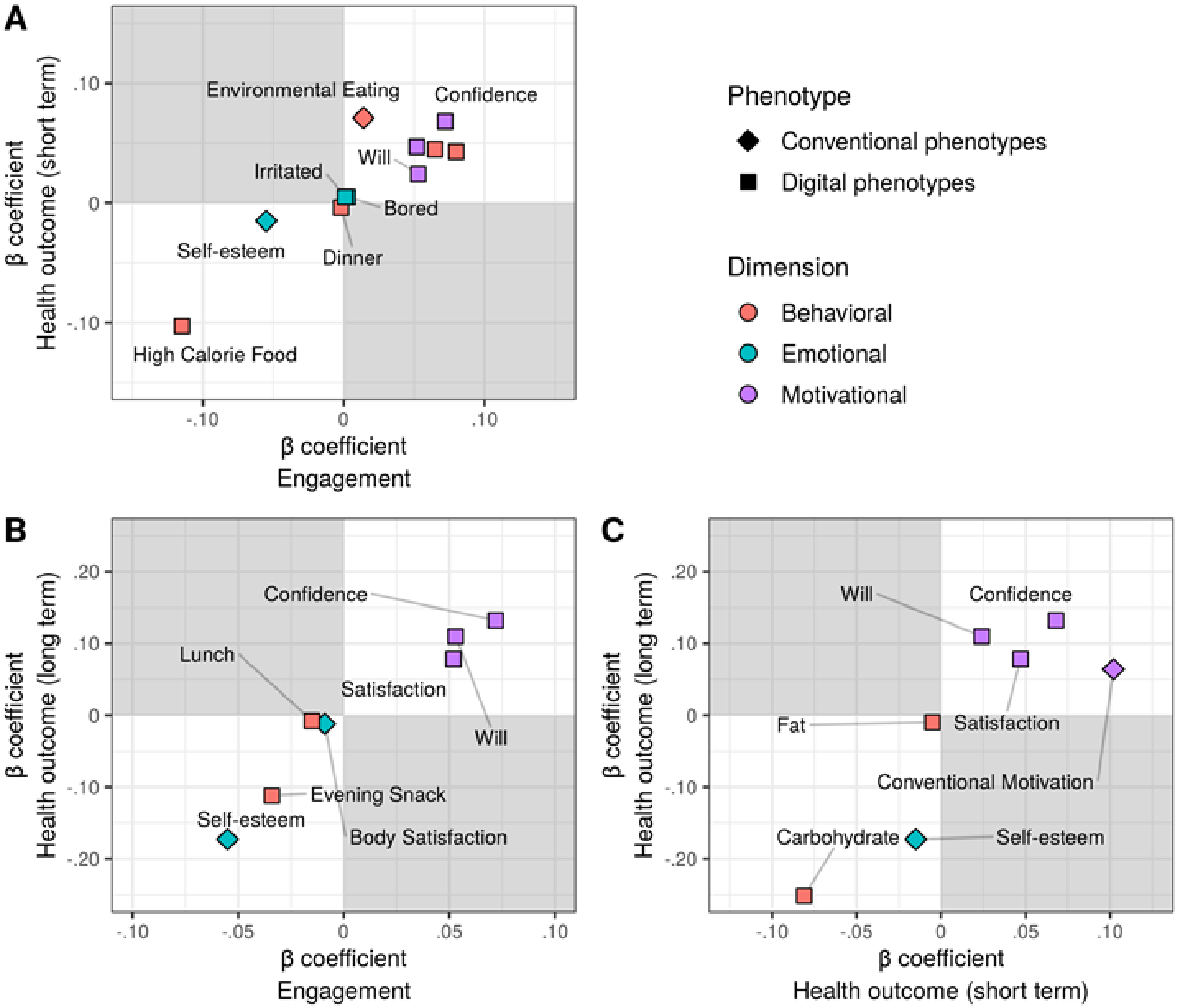


Table 2. Common and specific predictors of conventional and digital phenotypes for predicting engagement and health outcomes.

Phenotypes	Common predictors[a]	Predictors specific to each dependent variable		
		Engagement	Health outcome (short term)	Health outcome (long term)
Conventional phenotypes				
	• Self-esteem[b]	• Body satisfaction[b] • Environmental eating[c]	• Emotional eating[b] • Anxiety[b] • Environmental eating[c] • Conventional motivation[c]	• Food addiction[b] • Body satisfaction[b] • Conventional motivation[c] • Restrictive eating[c]
Digital phenotypes				
Behavioral				
	N/A[d]	• High-calorie food[b] • Night snack[b] • Lunch[b] • Dinner[b] • Breakfast[c] • Sugar[c] • Morning snack[c] • Moderate calorie food[c] • Low-calorie food[c] • Interaction frequency[c]	• High-calorie food[b] • Carbohydrate[b] • Sodium[b] • Fat[b] • Afternoon snack[b] • Low-calorie food[c] • Interaction frequency[c]	• Carbohydrate[b] • Night snack[b] • Lunch[b] • Fat[b] • Steps[b]
Emotional				
	N/A	• Irritated[c] • Bored[c] • Depressed[c]	• Irritated[c] • Bored[c]	N/A
Motivational				
	• Satisfaction[c] • Will[c] • Confidence[c]	N/A	N/A	N/A

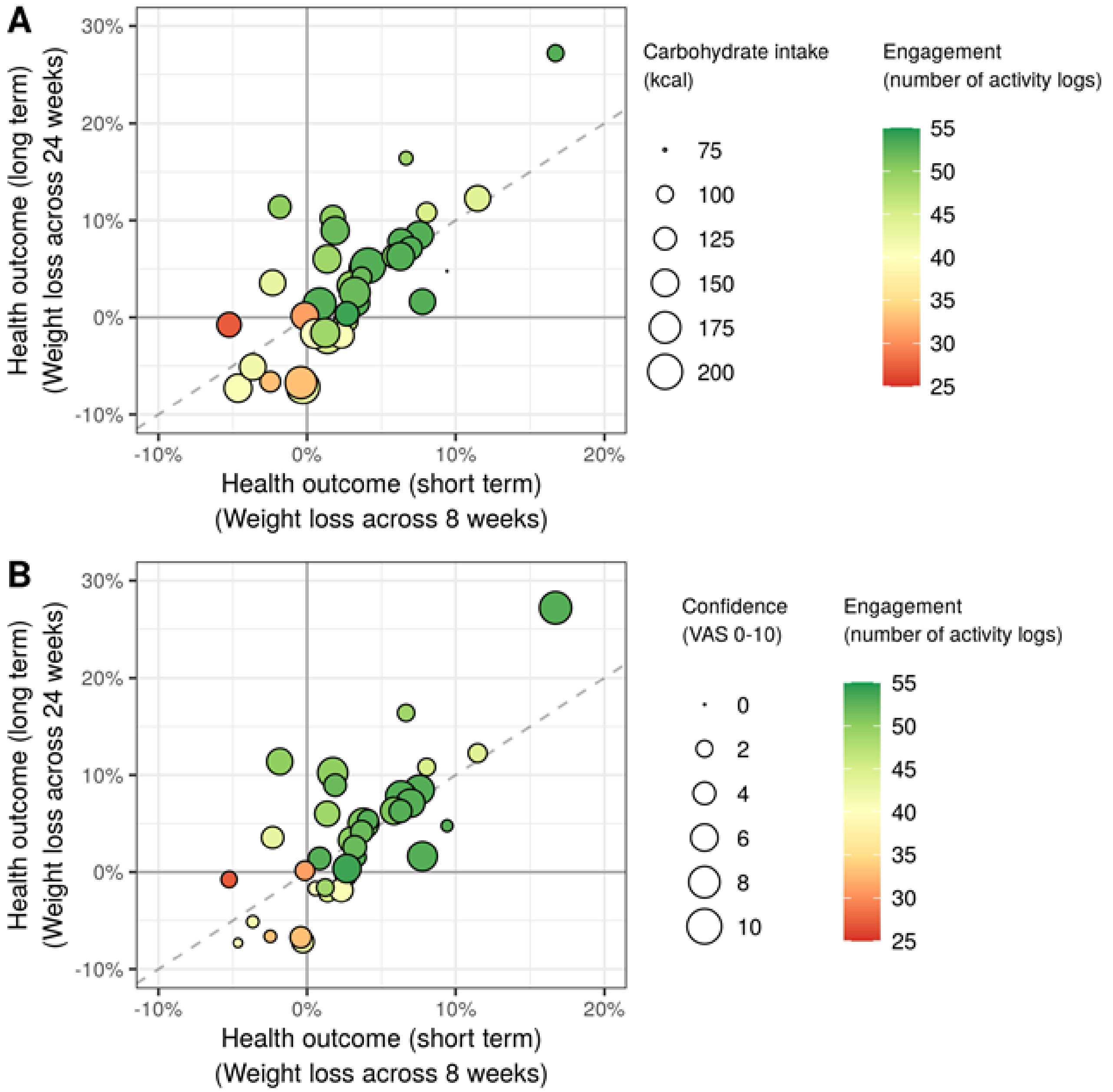
[a] Common predictors in the first column were involved in all models. The cognitive dimension of digital phenotypes was omitted because of a lack of significance.

[b] Predictors having positive associations with the engagement in app or health outcomes.

[c] Predictors having negative associations with the engagement in app or health outcomes.

[d] N/A: not applicable.

Figure 5. Two examples of common predictors between short-term and long-term health outcomes: (A) carbohydrate intake and (B) confidence in digital phenotypes. VAS: Visual Analogue Scale.



Regarding the model performance of the three prediction models, the machine learning approaches successfully predicted the engagement rate (mean R^2 0.416, SD 0.006), short-term weight change (mean R^2 0.382, SD 0.015), and long-term weight change (mean R^2 0.590, SD 0.011). In predicting long-term weight change, approximately 59% of the outcome variance was explained by the prediction model. In summary, these model performances suggest that the multivariate profiles in conventional and digital phenotypes provide phenotypes that are significantly associated with engagement and health outcomes.

Discussion

Principal Findings

Using a machine learning approach based on elastic net regression, we successfully demonstrated the applicability of the conceptual paradigm with complex dimensions of how in-app engagement is formed and affects health outcomes. This study showed that mobile apps' engagement was significantly associated with health outcomes, even 4 months after the cessation of digital interventions. We also found that both conventional motivation (before the intervention) and in-app motivation (during the intervention) were closely related to both engagement and clinical outcomes. Multiple aspects of motivation before and during the intervention could be used to predict engagement and health outcomes. Furthermore, both engagement and health outcomes are associated with multivariate psychological indices patterns, such as behavioral, cognitive, emotional, and motivational components, driven by regularized multivariate profiles obtained using the machine learning approach. From the results, we conclude that individuals' psychological states are the primary elements that influence engagement and health outcomes.

This study makes a clear implication on how engagement with apps influences clinical outcomes. Our finding that a higher frequency of logging into an app drives more significant improvements in health outcomes during the active intervention period is consistent with previous studies [27,28]. However, a notable finding in this study is that those who logged into the app more frequently also showed more favorable health outcomes after the cessation of the active intervention period. These results indicate that engagement is paramount to the app's potential effectiveness for behavior change, leading to a change in symptomatology. Thus, it is feasible for clinicians and users to predict their health outcomes according to the intensity of their participation in apps.

Digital interventions via apps are not the only realm in which engagement is an issue. Both face-to-face and digital interfaces encounter difficulties in maintaining adherence and engagement with monitoring, medications, and psychotherapies [29]. Digital therapeutics are beneficial for monitoring and analyzing real-time data and reaching out to users without barriers in space and time; however, they are more applicable to offer immediate feedback and prevent attrition than face-to-face clinics. From this perspective, a previous meta-analysis claimed that integrating a human factor into the treatment is an actionable strategy to alleviate dropout rates in the digital intervention [30]. Our result is also supportive in that the number of messages (interaction frequency between the user and therapist) showed the highest positive standardized coefficient with engagement with the app. Taken together, we suggest that human feedback is involved in the development of digital therapeutics to strengthen the engagement rate, leading to greater clinical efficacy.

For the first time, this study evaluated the multiple dimensions of motivation at two different periods: before (conventional motivation) and during (in-app motivation) the intervention [31,32]. Previous studies assessed motivation at several time points but only one dimension (ie, usability or satisfaction with digital intervention) [33,34]. Furthermore, other studies measured multiple dimensions of motivation (ie, satisfaction, acceptability, and usability) but only assessed one period (ie, after the digital intervention) [35,36]. These previous designs have limitations in reflecting the users' true motivation and predicting both engagement and clinical outcomes. According to our results, the common predictors of both engagement rate and health outcomes were in-app motivational phenotypes, referred to as satisfaction with the intervention, desire to improve health outcomes, and self-confidence. The level of self-esteem at baseline was also a common predictor of both engagement and health outcomes. Moreover, before implementing the intervention, the level of motivation was strongly related to health outcomes in both the short- and long-term courses. Altogether, these results suggest that motivation is the main component that determines engagement and health outcomes.

Previously, pragmatic qualities, systematic flow, satisfaction, usability, and esthetics were known as the major contributors to digital therapeutic engagement [7,27,29]. These previous results only serve as a basis for preliminary hypotheses on what may force engagement with apps. Few studies have examined engagement based on individuals' interactions with various intervention elements such as frequency of access, an average of steps, article views, and message views [28,37,38]. However, it is still challenging to establish a standardized approach to assess the engagement of these phenotypes because of various factors, such as diverse technological aspects, different intervention exposure times, and individual characteristics. Thus, we suggest measuring the multiple aspects of motivation directly before and during the intervention to predict dropout and give each participant individualized attention.

This is the first study to categorize diverse digital phenotypes into four different constructs: behavior, cognition, emotion, and motivation. This allows a comprehensive understanding of the nature of behavior change, which is closely related to the engagement and clinical outcomes of digital interventions. We suggest that the behavioral phenotypes (calorie density of food, snack time of the day, amount of food intake per meal, and frequency of message interactions with the therapists), emotional phenotypes (irritated, bored, and depressed), and motivational phenotypes (satisfaction, will, and confidence) are the favorable phenotypes for predicting the engagement in app and health outcomes. However, none of the cognitive phenotypes were capable of engaging in the app. The phenotypes predicting the health outcomes were similar but not identical to the engagement because the amount of nutritional intake was included instead of the amount of food intake per meal for the behavioral phenotypes, and depressive moods were excluded from the emotional phenotypes. These findings imply that not only users' physical participation in a specific target behavior (eg, logging food diary and number of steps) and behavior in digital spaces (eg, number of accesses) but also the user's psychological conditions (eg, emotion and motivation) are relevant to engagement and clinical outcomes.

To the best of our knowledge, this is the first study to apply a machine learning approach to provide relevant insights into improving both the adherence and clinical outcomes of digital interventions. Although previous mHealth intervention studies have shown that user engagement is critical to clinical outcomes, little effort has been made to conceptualize and estimate it. The major reason is that only a few mHealth programs predominantly use the applicable data to investigate participants' engagement or to examine its correlation with primary outcomes. However, we demonstrated the whole framework of how different types of phenotypes at baseline and during the intervention carry out in-app engagement and health outcomes. We used machine learning strategies with digital phenotypes to find an applicable model to predict intervention adherence for the first time.

This is also the first study to examine the determinants of significant weight changes from digital interventions. In addition, our first attempt to explore the phenotypes in two different periods (at baseline and during the intervention) and categorize them into four distinctive dimensions (behavior, cognition, emotion, and motivation) presents more comprehensive perceptions of engagement mechanisms and clinical outcomes. Finally, this study applied two specific methods, in-app and a web-based survey, for the first time to collect sufficient data, which led us to explore various components attaining favorable solutions for the issue of engagement and clinical efficacy in digital therapeutics. Using digital phenotypes and enhancing our insight into them to promote management will involve refined approaches for choosing and investigating diverse digital health data streams in a definite manner.

Limitations

This study had several limitations. First, all participants received cognitive behavioral therapy, so it lacked a control group that did not receive any intervention. Second, the number of participants was relatively small, which might not be sufficient for a reliable interpretation. However, as we extracted multivariate profiles to predict engagement and health outcomes, we remedied the shortage by using a machine learning approach. Furthermore, as this study explores the challenging concept of digital interventions, a small number of participants are still tolerable to apply the machine learning analysis [39]. Third, considering the relatively small sample size, the leave-one-out cross-validation may be sensitive to outliers in the data set. Furthermore, our study is somewhat exploratory, limited by the small sample size, which requires further investigation with large data sets to consolidate the validity of our findings. Finally, the experiment did not track longitudinal changes in health outcomes in the app.

Conclusions

Using a machine learning approach, we successfully established and validated an intuitive analytic strategy and provided visualization with a multiplex component paradigm of causality underlying digital psychotherapy on health outcomes. Our results revealed a key mechanism of psychological features interacting with multiple dimensions of motivation, which induce engagement in the app and enhance clinical efficacy. We expect that this study will play a significant role in establishing the most practical and effective mHealth intervention model, a vital insight for precision digital medicine.

Acknowledgments

The authors would like to thank Youngin Kim, the chief executive manager of Noom, for supporting this study by sharing the company’s advanced digital technologies. This study was supported by a grant from the National Research Foundation (NRF) of Korea, funded by the Korean government, Ministry of Science and Information and Communication Technologies (NRF-2018R1A5A2025964 to HJC and NRF-2018R1C1B3007313, NRF-2018R1A4A1025891, and BK21 FOUR Program [5199990314123] to WYA), the Institute for Information and Communications Technology Planning and Evaluation grant funded by the Korean government (Ministry of Science and Information and Communication Technologies; 2019-0-01367, BabyMind, to WYA), and the Creative-Pioneering Researchers Program through Seoul National University (WYA). The data that support the findings of this study are available from the authors’ GitHub repository [40].

Authors' Contributions

MK and JY contributed equally to this work. WYA and HJC are co-corresponding authors of this work. MK led and designed the conceptual framework of mHealth components. JY conducted the data analysis and performed a statistical or machine learning analysis. HJC conceived the project and provided valuable insights for devising and visualizing the concept of this project. WYA supervised the data analysis and served as a technical advisor for this project. MK and JY wrote the manuscript and edited the manuscript. HJC and WYA contributed to manuscript writing. All authors approved the final version of the manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Cross-validation results on the model performance of three elastic net models based on different mixing parameter values (α), the categorization of digital phenotypes and items used for each phenotype, and β estimates for conventional and digital phenotypes on the predictions of engagement and health outcomes.

[DOCX File , 234 KB - jmir_v23i6e27218_app1.docx]

Multimedia Appendix 2

CONSORT-EHEALTH (V 1.6.1) - Checklist.

[PDF File (Adobe PDF File), 938 KB - jmir_v23i6e27218_app2.pdf]

References

1. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]
2. Gunn JF, Lester D. Using google searches on the internet to monitor suicidal behavior. J Affect Disord 2013 Jun;148(2-3):411-412. [doi: 10.1016/j.jad.2012.11.004] [Medline: 23182592]
3. 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]
4. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
5. Tolou-Shams M, Yonek J, Galbraith K, Bath E. Text Messaging to Enhance Behavioral Health Treatment Engagement Among Justice-Involved Youth: Qualitative and User Testing Study. JMIR Mhealth Uhealth 2019 Apr 5;7(4):e10904 [FREE Full text] [doi: 10.2196/10904] [Medline: 30950808]
6. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955 [FREE Full text] [doi: 10.1002/asi.20801]
7. Wiebe E, Lamb A, Hardy M, Sharek D. Measuring engagement in video game-based environments: investigation of the user engagement scale. Comput Hum Behav 2014 Mar;32(8):123-132 [FREE Full text] [doi: 10.1016/j.chb.2013.12.001]
8. Przybylski AR, Rigby CS, Ryan RM. A motivational model of video game engagement. Rev Gen Psychol 2010 Jun 1;14(2):154-166 [FREE Full text] [doi: 10.1037/a0019440]
9. Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, et al. Seeing the "Big" picture: big data methods for exploring relationships between usage, language, and outcome in internet intervention data. J Med Internet Res 2016 Aug 31;18(8):e241 [FREE Full text] [doi: 10.2196/jmir.5725] [Medline: 27580524]
10. Al-Asadi AM, Klein B, Meyer D. Posttreatment attrition and its predictors, attrition bias, and treatment efficacy of the anxiety online programs. J Med Internet Res 2014 Oct 14;16(10):e232 [FREE Full text] [doi: 10.2196/jmir.3513] [Medline: 25316533]
11. Marsch LA. Opportunities and needs in digital phenotyping. Neuropsychopharmacology 2018 Jul;43(8):1637-1638 [FREE Full text] [doi: 10.1038/s41386-018-0051-7] [Medline: 29703995]
12. Ben-Zeev D, Scherer EA, Brian RM, Mistler LA, Campbell AT, Wang R. Use of multimodal technology to identify digital correlates of violence among inpatients with serious mental illness: a pilot study. Psychiatr Serv 2017 Oct 1;68(10):1088-1092 [FREE Full text] [doi: 10.1176/appi.ps.201700077] [Medline: 28669285]
13. Martinez-Martin NI, Insel TR, Dagum P, Greely HT, Cho MK. Data mining for health: staking out the ethical territory of digital phenotyping. NPJ Digit Med 2018;1:1 [FREE Full text] [doi: 10.1038/s41746-018-0075-8] [Medline: 31211249]
14. Lutz WS, Schwartz B, Hofmann SG, Fisher AJ, Husen K, Rubel JA. Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: a methodological proof-of-concept study. Sci Rep 2018 May 18;8(1):7819 [FREE Full text] [doi: 10.1038/s41598-018-25953-0] [Medline: 29777110]
15. Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton D, Clifford GD. Machine learning and decision support in critical care. Proc IEEE Inst Electr Electron Eng 2016 Feb;104(2):444-466 [FREE Full text] [doi: 10.1109/JPROC.2015.2501978] [Medline: 27765959]
16. Kim M, Kim Y, Go Y, Lee S, Na M, Lee Y, et al. Multidimensional cognitive behavioral therapy for obesity applied by psychologists using a digital platform: open-label randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 30;8(4):e14817 [FREE Full text] [doi: 10.2196/14817] [Medline: 32352391]
17. Cooper Z, Hawker DM, Fairburn CG. Cognitive-Behavioral Treatment of Obesity: A Clinician's Guide. New York City: The Guilford Press; 2003.
18. Juresa Z, Musil V, Majer M, Ivanković D, Petrović D. Behavioral pattern of overweight and obese school children. Coll Antropol 2012 Jan;36 Suppl 1:139-146. [Medline: 22338763]
19. Jansen A, Houben K, Roefs A. A cognitive profile of obesity and its translation into new interventions. Front Psychol 2015;6:1807 [FREE Full text] [doi: 10.3389/fpsyg.2015.01807] [Medline: 26640451]
20. Jorm A, Korten A, Christensen H, Jacomb P, Rodgers B, Parslow R. Association of obesity with anxiety, depression and emotional well-being: a community survey. Aust N Z J Public Health 2003;27(4):434-440. [doi: 10.1111/j.1467-842x.2003.tb00423.x] [Medline: 14705308]
21. Guay F, Vallerand R, Blanchard C. On the assessment of situational intrinsic and extrinsic motivation: the situational motivation scale (SIMS). Motiv Emot 2000 Sep;24(3):175-213. [doi: 10.1023/A:1005614228250]
22. Sahib AS. Eating behavior in a sample of overweight and obese: a cross sectional study. J Obes Weight-Loss Medic 2016 Dec 31;2(2):-. [doi: 10.23937/2572-4010.1510014]
23. Kral TV, Moore RH, Chittams J, Jones E, O'Malley L, Fisher JO. Identifying behavioral phenotypes for childhood obesity. Appetite 2018 Aug 1;127:87-96 [FREE Full text] [doi: 10.1016/j.appet.2018.04.021] [Medline: 29709528]
24. Zou H, Hastie T. Regularization and variable selection via the elastic net. J Royal Statistical Soc B 2005 Apr;67(2):301-320 [FREE Full text] [doi: 10.1111/j.1467-9868.2005.00503.x]

25. Ahn W, Vassileva J. Machine-learning identifies substance-specific behavioral markers for opiate and stimulant dependence. Drug Alcohol Depend 2016 Apr 1;161:247-257 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.02.008] [Medline: 26905209]
26. Ahn W, Ramesh D, Moeller FG, Vassileva J. Utility of machine-learning approaches to identify behavioral markers for substance use disorders: impulsivity dimensions as predictors of current cocaine dependence. Front Psych 2016 Mar 10;7:34 [FREE Full text] [doi: 10.3389/fpsyt.2016.00034] [Medline: 27014100]
27. Graham AK, Lattie EG, Mohr DC. Experimental therapeutics for digital mental health. JAMA Psychiatry 2019 Dec 1;76(12):1223-1224 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.2075] [Medline: 31433448]
28. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing user engagement of an mhealth intervention: development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
29. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 5;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
30. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
31. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 1;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
32. Borghouts J, Eikey E, Mark G, de Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
33. Bauer AM, Hodsdon S, Bechtel JM, Fortney JC. Applying the principles for digital development: case study of a smartphone app to support collaborative care for rural patients with posttraumatic stress disorder or bipolar disorder. J Med Internet Res 2018 Jun 6;20(6):e10048 [FREE Full text] [doi: 10.2196/10048] [Medline: 29875085]
34. Ben-Zeev D, Brian RM, Jonathan G, Razzano L, Pashka N, Carpenter-Song E, et al. Mobile health (mHealth) versus clinic-based group intervention for people with serious mental illness: a randomized controlled trial. Psychiatr Serv 2018 Sep 1;69(9):978-985. [doi: 10.1176/appi.ps.201800063] [Medline: 29793397]
35. Dahne J, Kustanowitz J, Lejuez CW. Development and preliminary feasibility study of a brief behavioral activation mobile application (behavioral apptivation) to be used in conjunction with ongoing therapy. Cogn Behav Pract 2018 Feb;25(1):44-56 [FREE Full text] [doi: 10.1016/j.cbpra.2017.05.004] [Medline: 29725198]
36. Hidalgo-Mazzei D, Mateu A, Reinares M, Murru A, del Mar CB, Varo C, et al. Psychoeducation in bipolar disorder with a SIMPLe smartphone application: feasibility, acceptability and satisfaction. J Affect Disord 2016 Aug;200:58-66. [doi: 10.1016/j.jad.2016.04.042] [Medline: 27128358]
37. Baltierra NB, Muessig KE, Pike EC, le Grand S, Bull SS, Hightow-Weidman LB. More than just tracking time: complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016 Feb;59:299-307 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]
38. Davies C, Corry K, van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]
39. Tazawa Y, Liang K, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, et al. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. Heliyon 2020 Feb;6(2):e03274 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03274] [Medline: 32055728]
40. Yang J. Digital phenotypes and mobile health (2020). GitHub. URL: https://github.com/CCS-Lab/digital-phenotypes-mHealth_2020 [accessed 2021-06-09]

Abbreviations

BSQ: Body Shape Questionnaire
CDI: calorie density index
dCBT: digital cognitive behavioral therapy
mHealth: mobile health
NRF: National Research Foundation
RMSE: root mean squared error
SIMS: Situational Motivation Scale

Edited by R Kukafka; submitted 16.01.21; peer-reviewed by C Carrion, H Ranjani, L Guo; comments to author 20.03.21; revised version received 28.04.21; accepted 06.05.21; published 24.06.21.

Please cite as:
Kim M, Yang J, Ahn WY, Choi HJ
Machine Learning Analysis to Identify Digital Behavioral Phenotypes for Engagement and Health Outcome Efficacy of an mHealth Intervention for Obesity: Randomized Controlled Trial
J Med Internet Res 2021;23(6):e27218
URL: https://www.jmir.org/2021/6/e27218/
doi:10.2196/27218
PMID:34184991

Original Paper

Mobile-Based and Cloud-Based System for Self-management of People With Type 2 Diabetes: Development and Usability Evaluation

Raheleh Salari[1], PhD; Sharareh R Niakan Kalhori[1], PhD; Marjan GhaziSaeedi[1], PhD; Marjan Jeddi[2], MD, PhD; Mahin Nazari[3], PhD; Farhad Fatehi[4,5], MD, PhD

[1] Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran
[2] Endocrinology and Metabolism Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[3] Department of Health Promotion and Education, School of Health, Shiraz University of Medical Sciences, Shiraz, Iran
[4] School of Psychological Sciences, Monash University, Melbourne, Australia
[5] Centre for Health Services Research, The University of Queensland, Brisbane, Australia

Corresponding Author:
Sharareh R Niakan Kalhori, PhD
Department of Health Information Management
School of Allied Medical Sciences
Tehran University of Medical Sciences
No17, FareDanesh Alley, Oods St.
Eghelab Ave
Tehran, 1417653761
Iran
Phone: 98 2188982886
Fax: 98 2188983017
Email: sh-rniakank@sina.tums.ac.ir

Abstract

Background: As the use of smartphones and mobile apps is increasing, mobile health (mHealth) can be used as a cost-effective option to provide behavioral interventions aimed at educating and promoting self-management for chronic diseases such as diabetes. Although many mobile software apps have been developed for this purpose, they usually lack a theoretical foundation and do not follow the guidelines suggested for evidence-based practice. Therefore, this study aimed to develop a theory-based self-management app for people with type 2 diabetes and provide an app based on a needs assessment analysis.

Objective: This paper describes the development and usability evaluation of a cloud-based and mobile-based diabetes self-management app designed to help people with diabetes change their health behavior and also enable remote monitoring by health care providers.

Methods: The development of this mHealth solution comprises 3 phases. Phase I: feature extraction of the Android apps that had a user rating of 4 stars or more and review of papers related to mHealth for diabetes self-management were performed followed by seeking expert opinions about the extracted features to determine the essential features of the app. Phase II: design and implementation included selecting which behavioral change and structural theories were to be applied the app and design of the website. Phase III: evaluation of the usability and user experience of the mobile app by people with diabetes and the portal by health care providers using the User Experience Questionnaire.

Results: The developed mobile app includes modules that support several features. A person's data were entered or collected and viewed in the form of graphs and tables. The theoretical foundation of behavioral intervention is the transtheoretical model. Users were able to receive customized messages based on the behavioral change preparation stage using the Kreuter algorithm. The clinician's portal was used by health care providers to monitor the patients. The results of the usability evaluation revealed overall user satisfaction with the app.

Conclusions: Mobile- and cloud-based systems may be an effective tool for facilitating the modification of self-management of chronic care. The results of this study showed that the usability of mobile- and cloud-based systems can be satisfactory and promising. Given that the study used a behavioral model, assessment of the effectiveness of behavior change over time requires further research with long-term follow-up.

(J Med Internet Res 2021;23(6):e18167) doi:10.2196/18167

KEYWORDS

type 2 diabetes; mobile health; mHealth; mobile app, self-management; behavior change

Introduction

Diabetes is a chronic condition in which the pancreas is unable to produce enough insulin to regulate glucose or the body cells cannot respond to insulin properly [1]. Diabetes is one of the most common metabolic causes of mortality because of its complications [2]. The prevalence of diabetes worldwide was estimated at 8.4% in 2017. This is projected to increase by 1.5% in 2045 [3]. The high prevalence of diabetes has high social and financial consequences, especially in low- and middle-income countries [3], including the Middle East and North Africa region [4]. The prevalence of diabetes in Iran is high and the number of affected people is continuously increasing. From 2005 to 2011, the prevalence of diabetes has grown by 35% in Iran [5-7]. The frequencies of diabetes-related complications among people with diabetes admitted to tertiary care centers in Iran are relatively high [7].

To minimize the potential risks of diabetes-related complications, patients should be educated and monitored to enhance their self-management abilities. People with diabetes who are empowered by self-management abilities show improvements in health outcomes [8]. People with diabetes must have access to ongoing health care services. However, in developing countries, financial and human resources are limited. Face-to-face education and self-management training for a person's empowerment to control their illness are often suboptimal and limited [1]. Due to limited resources, the use of information technology has been suggested to improve a person's self-management skills regularly. Although the use of the latest mobile technologies is valuable in empowering people with diabetes for self-management, health-related mobile apps are still relatively new and need more attention in developing countries [1].

The popularity of mobile health technologies has become an opportunity for education, remote monitoring, self-management, and data collection for diabetes care [9]. This opportunity can be used to provide interventions for user-centered and evidence-based self-management for people with diabetes [10]. Meanwhile, the development of these technologies, referred to as mHealth, has been greater than other apps [9]. mHealth is defined as "medical and public health measures supported by mobile devices, such as cell phones, personal monitoring devices, personal digital assistants, and other wireless devices" [11]. We have seen a large number of developed apps in the clinical care setting for patient self-management intervention [12], but few apps developed based on validated behavioral theories [13]. Interventions will have a greater chance of changing a person's behavior if they use a behavioral model than interventions that have not used any behavioral model [14].

It is noteworthy that each behavioral theory can help us to understand why people behave concerning their health. Therefore, approaches that do not use these theories might fail [15]. Behavioral theory-based data collection for diabetes care can lead to customized feedback to enhance self-management skills. This feedback is a fundamental aspect of changing behavior that leads to improvement in a person's self-management [9]. However, there is still no definite opinion on which behavioral models should be used to design an appropriate intervention for improving diabetes care [16]. Research shows that cognitive theories such as theory of planned behavior; transtheoretical model (TTM); and use of self-efficacy, information motivation, and behavioral skills were most widely used in technology-based interventions, especially mHealth apps [14].

One particular model provided by Prochaska and DiClemente [17] is the transtheoretical model. In this model, behavior change is considered as a 5-step process over time rather than an event. Since the results obtained from the use of this model have been reported to be useful for patients with type 2 diabetes mellitus (T2DM) [18], we decided to use this model in our study.

Two important points motivated the conduct of this study. In contrast to developed countries, in Iran, with the highest rate of diabetes prevalence in the Middle East and North Africa region [19,20], no study has been conducted to design a cloud- and mobile-based intervention underpinned by a behavioral model. Second, while most of the apps have been developed in the English language and there are already extensive studies on the usability of such apps [13], language remains one of the barriers in the adoption of English mobile apps by patients in non-English speaking countries [21]. There are numerous English apps in the public app stores for T2DM. Despite a large number of apps in this field, few of them are supported by scientific evidence, have been designed based on a behavioral model, and endorsed by health care professionals.

We felt the need for a Persian app for Iranian patients. However, we were not able to find any app in the Persian language that can provide management, monitoring, and education modules to Iranian patients based on a behavioral model. This paper reports on first Persian mobile app for diabetes care developed in Iran. In this study, we aimed to create a cloud- and mobile-based system for people with T2DM and health care providers to support diabetes self-management. We also evaluated the usability of the app. This paper sets out the steps taken to design and implement this app and highlights the limitations and results of the study. We believe the insight gained in this study illuminates the path for future studies in this field.

Methods

Study Overview

This study consisted of three main phases. To achieve a theory-based mobile- and cloud-based system that delivers a set of tailored messages, we first had to determine and define the requirements of such a system. One of these requirements was a behavioral model as a basis to design appropriate messages. Therefore, in the next phase, we determined which behavioral model we wanted to use according to the intended purpose of the system. The choice of model should be such that it can have the capacity to track patients over time and provide a suitable strategy for each stage of changing the patient's care process. After that, the third phase was to design the mobile app and cloud-based system that included modules to meet the suggested requirements.

Phase I: Needs Assessment and Planning

Initially, we conducted a survey to determine what features and functions are required for an ideal mobile app for people with T2DM [22]. We reviewed the literature and available diabetes apps for iOS and Android to investigate the features of diabetes apps to get familiar with them. The inclusion criteria were free apps directly related to diabetes with 4 or more stars (out of 5) in user ratings. Exclusion criteria were supporting only a single feature (for example, insulin calculation, recording data in an electronic notepad); not designed for diabetes self-management; provide information only such as how to use glucometer or educational materials about diabetes; not updated within 12 months prior to the search date; or solely targeting fitness, physical activity, or diet of people with diabetes. Relevant English articles indexed in Web of Knowledge, Google Scholar, PubMed, Scopus, and Science Direct published from 2012 to 2017 were retrieved using the keywords "diabetes," "glucose," "blood sugar," "insulin," "mobile Health," "mobile apps," "smartphone," "mobile phone," and "mHealth." Articles that did not directly mention the use of mobile devices for the self-management of T2DM were excluded.

The validity of extracted features was then analyzed. For this task, we examined both the relevancy and necessity of features. We requested the members of 6 national interdisciplinary expert teams, including the Iranian Board of Health Informatics, Health Information Management, Endocrinology, and Health Education and Promotion, to provide comments on the most necessary and relevant features of a mobile app for diabetes management. Afterward, we designed a Likert-style questionnaire by verified features, to determine the final features list, and analyzed their level of importance (Multimedia Appendix 1). The response options to the questions ranged from 1=totally disagree to 5=totally agree. A total of 21 experts participated in the survey. The details of this phase have been reported elsewhere [22].

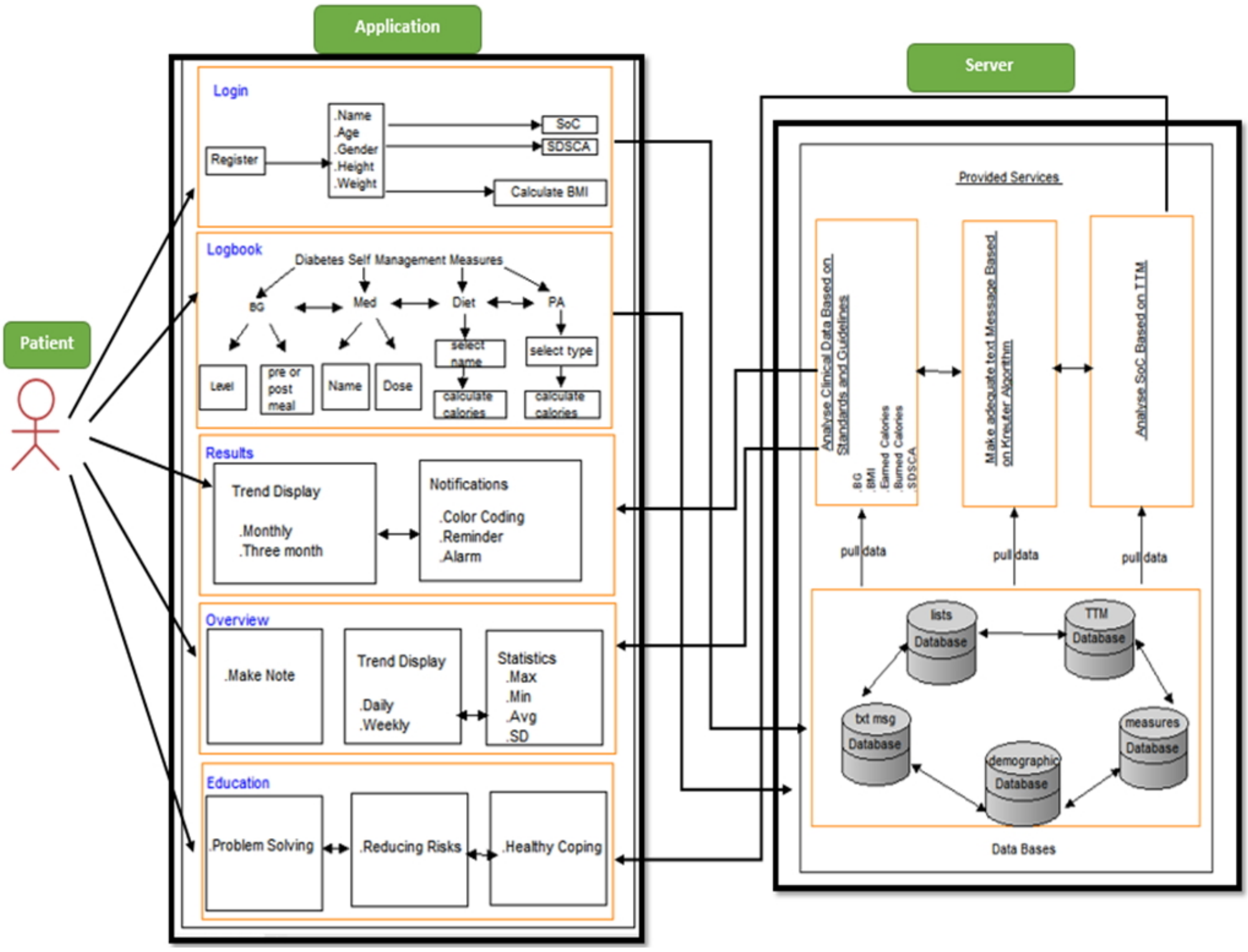
Phase II: Design and Implementation

System Architecture and Model

A 3-tier model was applied as the conceptual model of the app. At the highest level, the logic layer presents the main functions of the app. This layer is responsible for processing the data and rendering them to the display layer. The data layer provides an interface to the logic layer and performs the necessary operations, including storing, editing, deleting, and retrieving data without engaging in the complexity of the database. In this layer, the database is designed and used. We needed to use a suitable model for customization of the message here. In the display layer, the application is placed under a web browser and a mobile app.

The mobile apps are designed for people with T2DM and health care providers. The users record the required data using the app and connected devices. Based on the status of the patient's behavioral stage and other caregiving data such as blood glucose, physical activity, and calorie intake, message embedded in the library on the application server was called and displayed on the mobile phone screen. It is also possible to view care charts of the user at weekly, monthly, quarterly, and yearly intervals. The clinicians, on the other side, can monitor the patients' data and be informed of their medication status, blood sugar, nutrition, and physical activity. The clinicians also can view notifications and messages sent to the patients. Figures 1 and 2 provide an overview of the framework designed for this cloud-based system and the mobile app.

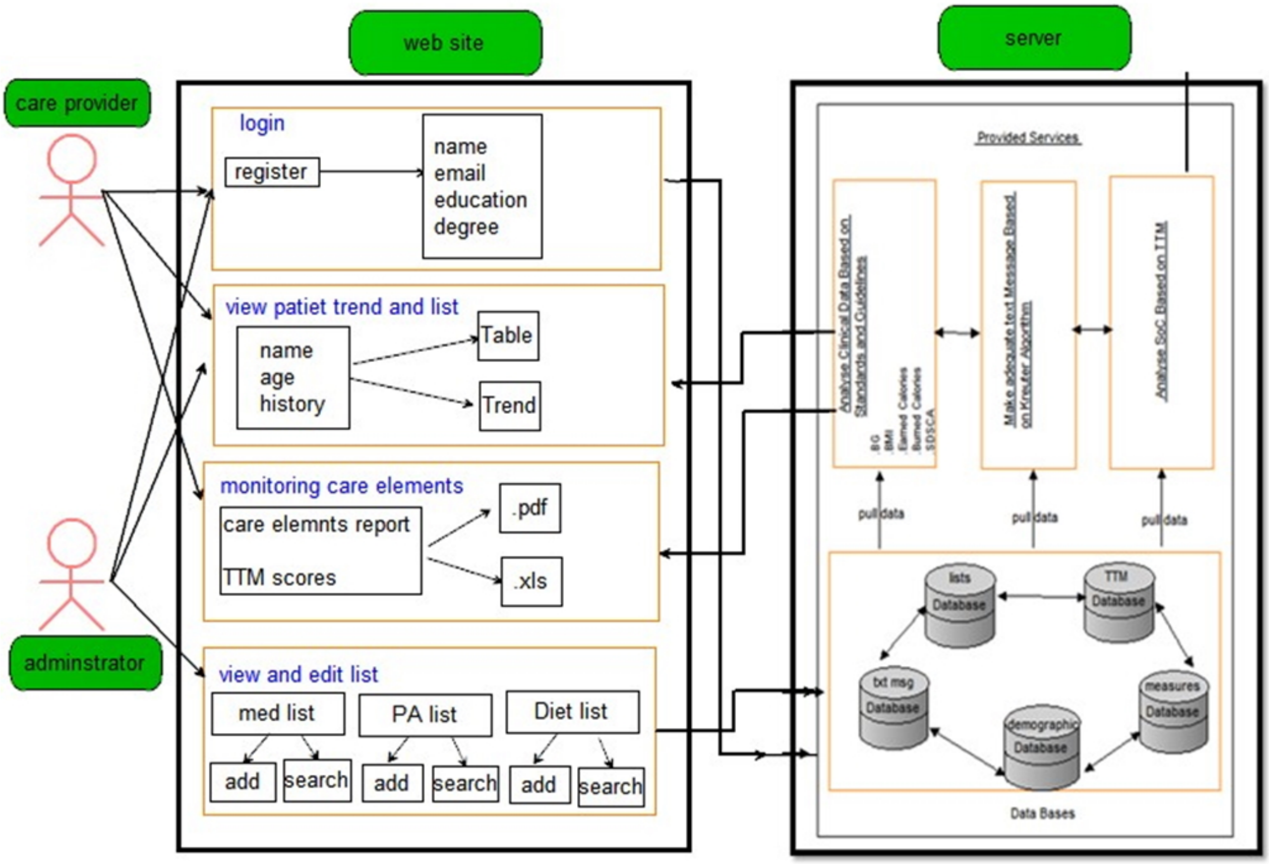
Figure 1. Proposed architectural framework for functionalities of a system for people with diabetes.



The self-management app was designed to enable people with diabetes to manage their required care by monitoring blood glucose, physical activity, and diet as well as leverage the behavioral stage by receiving customized messages using smartphones. The app includes 5 modules that provide a platform for facilitating diabetes management, as suggested by American Diabetes Association. These modules include log-in and data management, logbook, analysis, overview, and education [23]. By module we mean a group of related features for performing a function in a mobile app. In this way, each module includes a set of features with specific tasks. For example, the overview module can provide an overview of the results of the care parameters over different periods, such as weeks, months, and years.

Many different theories guide health interventions. We explored and compared the existing behavioral models to identify the best model for further adaptation of the system according to the research goals and needs. To do this, the existing models were reviewed to reveal which model would help people with T2DM the most, and can be used for customizing messages for behavior change at a specified stage. TTM was identified as the best one for this purpose and formed the theoretical base of the app. TTM considers a person's behavioral change in a chronic status as a multistep process [17]. In this case, the person's behavioral stage status can be determined from the beginning (time of registration in the service) to the end (conclusion of the service). Staging of individual behaviors can be performed in two ways. One approach is to identify key behaviors based on their importance in diabetes management. They then implement step-by-step interventions for behavior change. The second approach is to focus on behavior that is almost ready for change [24]. In this study, we use the first approach, focusing on the important self-management behaviors of diabetes and monitoring the stages of change. These behavior aspects include physical activity, blood glucose monitoring, diet and nutrition, and medication, according to the International Diabetes Federation recommendations [25].

Figure 2. Proposed architectural framework for functionalities of a system for health care providers.



We also used the Kreuter algorithm, which was proposed in 1999, as a structured process to determine the customization of interventions. The reason for choosing this algorithm is to determine the level of intervention presented in this study that, based on a behavioral model, aims to deliver customized messages to help the self-management ability of diabetes persons [26]. The first step of the Kreuter algorithm is instrument identification and analysis. At this stage, a behavioral model is usually used as the baseline model to identify items for better diabetes care. In this study, the basic behavioral model for identifying the person's problem is TTM. The second step of the algorithm is to use a tool or construct to evaluate the current condition of a person with T2DM by asking some questions about diet, medication, blood glucose measurement, and physical activity habits and behaviors. In the third step, messages were designed, and a library of messages was created. We chose those messages from publications that had developed the messages according to a scientific methodology. The messages were chosen according to their validity and relevance in the area of diabetes. To prepare a set of messages for this phase, we explored similar messages in past studies and existing educational resources and met several times with a health education specialist. We use the online training resources of the Iranian Ministry of Health and Medical Education [27], Association of Diabetes Care & Education Specialists [28], International Diabetes Federation [29], and American Diabetes Association [30]. In the fourth step of this algorithm, according to the previous step, the necessary responses were prepared and different modes were considered. In the final step of the algorithm, the rational rules for the relationship between the status and the appropriate message were made (ie, in other words, the relationship between the database and the library was designed).

System Modules

Before creating an app, it is necessary to provide a structure of each module to determine its functions and the relationship between these modules. To meet this need, we used diagrams to represent the structure and relationship of data elements and other components involved in each module and a scenario to illustrate the activities of that module. These diagrams are used during app analysis to identify requirements and illustrate how the app works. Actors in the mobile app and web application design include person with T2DM, specialist physicians, diabetes care nurses, and system administrators. Scenarios are the details of the set of commands or activities performed in a particular routine in an information system and include a sequence of operations that objects perform in the system.

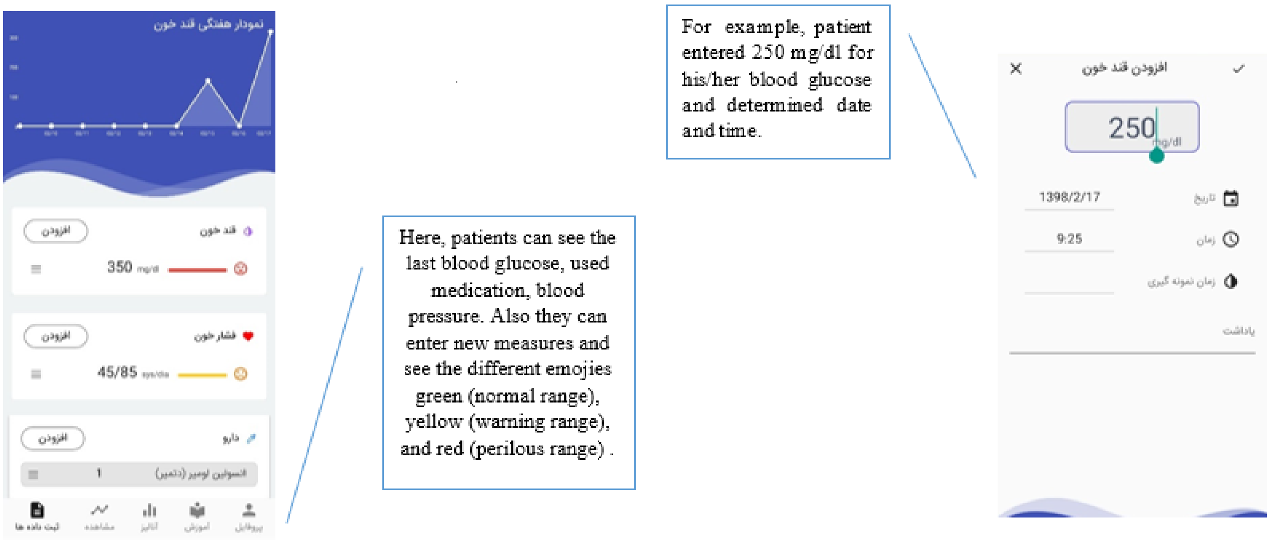
As illustrated in Figure 1, a person with diabetes communicates by smartphone with the server. In this figure, the person with diabetes is the end user of the app, which encompasses 5 data modules. In Figure 2, on the server side, the databases include demographic information, data collection and analysis, drug and food list information, message and stage information, and the person's stage according to TTM. The data for each database is called for its segment on which the relevant analysis is based. For example, a person's data in addition to their stage of changes based on TTM is invoked and combined with the messages in the relevant database and displayed to them.

On the user side, the health care provider and the webmaster are the end users of the web portal, each with different levels of access. Moderator modules include log-in, access to medication list, physical activity, and food names. The manager has access to the list of health care providers. The health care provider has access to a list of people with diabetes and their care process, as well as a score based on the level of change, received and reported by TTM. Patients are scored in order of each stage of the change. For example, a patient who is in the first step of the physical activity does not even think about having an exercise program and gets a score of 1, and a patient who is in the fifth step of physical activity has been exercising for more than 6 months. On the server side, databases include demographic information, measurement and analysis information, drug and food list information, message information, and stage information according to TTM. The data for each database is called for its segment on which the relevant analysis is based. For example, data on monitoring a person's information from their self-management process is called from the relevant database and displayed to the health care provider.

The use of the app begins with registration. The person enters the phone number (which triggers a confirmation code to be sent to the mobile number). The patient then enters a password and demographic information. A questionnaire based on TTM is presented in this section. The person completes the options for this questionnaire. The data from this section will be used to apply the rules of the Kreuter algorithm. People with diabetes should also include information about their weight, height, and level of physical activity. As mentioned earlier, these data will be used to calculate the number of calories needed and adjust their diet.

In the logbook module (Figure 3), the person's blood glucose levels, medications, diet, and physical activity should be entered. By selecting the 2-hour or fasting blood glucose (BG), the type of blood sugar measured will be recorded. In addition to receiving text messages, the app will display an emoticon that shows the patient's blood sugar status in green (normal range), yellow (warning range), and red (danger range). Another parameter is the list of medications used. These medications are listed in alphabetical order, and pills and insulin can be selected from this list and the date and method of administration recorded. The next parameter is the database of foods that can be continuously updated. The app includes customized Iranian food databases for dietary intake. When a user exceeds the number of calories recommended for a day compared with the normal level and required calories based on the basic data, the person will receive a notification informing them to observe their diet. The food list is arranged alphabetically and allows people to determine the amount of food consumed. Depending on the type and amount of food consumed, the person's intake calorie count will be calculated. For the physical activity, the person can select name and duration of their activities from the list, and their activity level will be calculated based on International Physical Activity Questionnaire.

Figure 3. Logbook module where participants can enter their clinical data. This screen shows that the blood glucose level was entered and saved on a specific day and time.



Here, patients can see the last blood glucose, used medication, blood pressure. Also they can enter new measures and see the different emojies green (normal range), yellow (warning range), and red (perilous range) .

For example, patient entered 250 mg/dl for his/her blood glucose and determined date and time.

قند خون 350 mg/dl

فشار خون 45/85 sys/dia

دارو

انسولین لومیر (دتمیر) 1

افزودن قند خون

250 mg/dl

تاریخ 1398/2/17

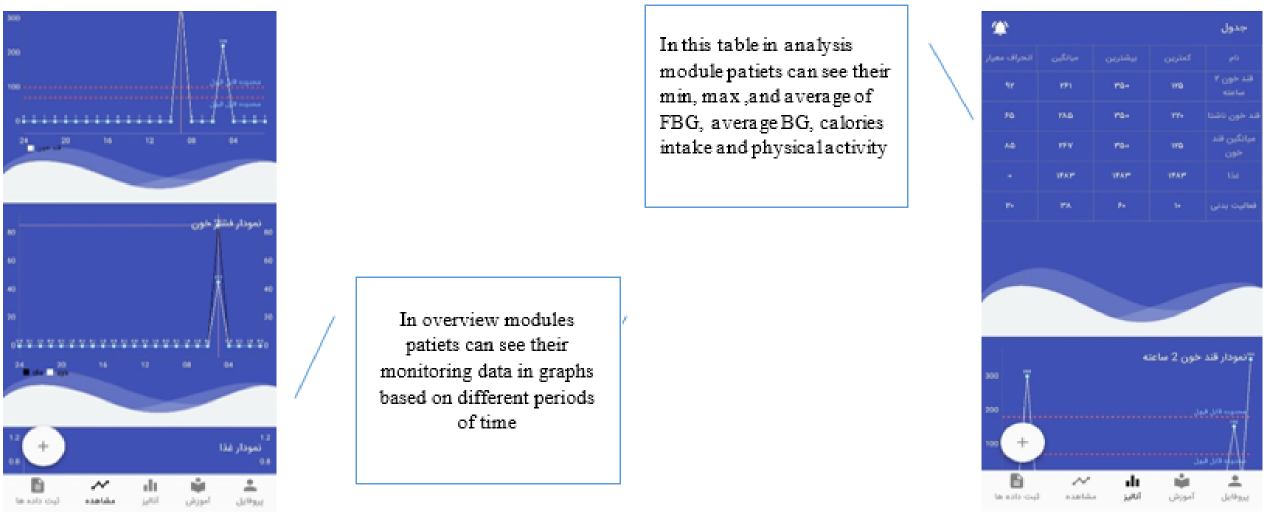
زمان 9:25

زمان نمونه گیری

یادداشت

In the overview and analysis modules (Figure 4), users can see the daily and weekly charts of the 4 self-management aspects. There is also a graph to compare the calorie intake versus the number of calories consumed. Another feature of this module is that person's data can be examined in a data table. In this table, we analyzed the statistical indices of mean, standard deviation, maximum and minimum blood sugar, calorie intake, and calories burned.

Figure 4. Overview and analysis module: participants can view the entered data in a chart format and see trends over time. Charts show physical activity range and calorie intake.



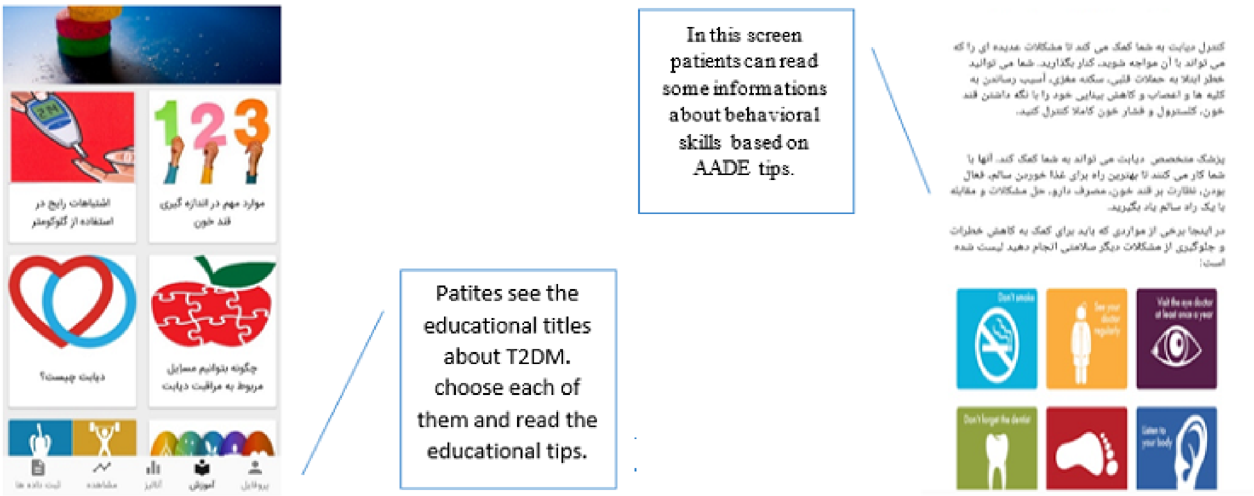
In this table in analysis module patiets can see their min, max ,and average of FBG, average BG, calories intake and physical activity

In overview modules patiets can see their monitoring data in graphs based on different periods of time

The data in the overview module can be explored in greater detail in the analysis module. In the analysis module of recorded data, monthly, quarterly, and annually surveyed comparisons are presented in linear graphs. The data comparison table of this module contains data in these time intervals. Another feature of this module is the ability of the people to define their own reminders. The user can set a specific time for a reminder during the process of self-management.

In the education module (Figure 5), users can receive general education and customized messages. The education module makes this information available to users regardless of the stage of change. The content of this module is also derived from the texts used to customized messages in addition to behavioral self-management recommendations, which are recommended by the Association of Diabetes Care & Education Specialists. This education includes tips on reducing risk, problem solving, healthy coping, and some general information about diabetes and using a glucometer.

Figure 5. Education module: participants will learn tips on controlling diabetes, improving problem-solving skills, reducing risks, and healthy coping.



In this screen patients can read some informations about behavioral skills based on AADE tips.

Patites see the educational titles about T2DM. choose each of them and read the educational tips.

For the health care providers, the web portal opens to a summary page that displays the person’s history and their trend in BG, diet and calories, physical activity, and behavioral stages. Selecting a person links to a web page displaying their contact details and details of their diabetes medication and expected number of BG recordings per week, month, and year. Once a person is selected, health care professionals can see a tabular display of BG readings that mimics a paper diary.

Graphical representations similar to the graphs displayed on the phone are also implemented on the web portal. The database, algorithms, and user interface for people with diabetes and health care providers were realized through coding. The developed app used the Android SDK platform 4.4.2 Java Development Kit. The database management program was developed using MYSQL. The app works on mobile phones running the Android operating system versions 2.3 to 4.4. The Laravel framework and the PHP programming language have been used to develop cloud-based software.

Phase III: Usability Evaluation

The usability of the system and user satisfaction were assessed using health care providers and diabetes person’s data via using the User Experience Questionnaire (UEQ). UEQ is a standardized questionnaire in which end users describe their perception regarding aspects such as whether the app is easy to use, clear, confusing, and so on. This questionnaire measures 6 scales: efficiency, perspicuity, attractiveness, dependability, simulation, and novelty. The scales of the UEQ cover a comprehensive impression of user experience. Both classical usability aspects (efficiency, perspicuity, dependability) and user experience aspects (originality, stimulation) are measured. It consists of 26 contrarian adjective pairs randomly ordered to represent the 6 scales. The items are scaled from –3 to +3 with –3 representing the most negative answer, 0 a neutral answer, and +3 the most positive answer. All 14 recruited users evaluated the app for 10 days. This questionnaire has a useful tool developed in Excel (Microsoft Corp) that interprets the results and compares them with the results of previous studies in the same field.

Participation was voluntary, and the respondents could opt out of any phase of the study at any time. All participants were fully informed about the project. The identities of the participants were kept confidential throughout the process of data collection. The UEQ questionnaire was used to evaluate the app in the section based on the smartphone for people with diabetes and a cloud-based app for health care providers. We recruited 14 people with diabetes at the beginning, and all of them concluded the study. People were included if they met predefined criteria: aged over 18 years and under 60 years, have a smartphone with Android OS versions 2.3 to 4.4, not be insulin-dependent, at least 2 years have passed since they developed T2DM, be literate about using a smartphone, and be willing to participate in the study. The usability evaluation was conducted in Shahid Motahari Clinic of Shiraz and lasted for 10 days. Due to the limited number of health care providers (n=7), we did not consider any criteria and invited those who were interested in participating in the study to use and evaluate the intervention.

Results

Verified Features

The details of methods and results of the literature review as well as the process of selecting and verifying the features were published previously [22]. Based on expert opinions, 23 relevant features out of 33 were approved. These features were included: blood glucose, insulin and medication, physical activity, diet, weight and BMI, and blood pressure tracking; mealtime tagging; food database; educational materials; healthy coping; reducing risks; problem solving; messaging; color coding; alerts; reminders; target range setting; trend chart view; logbook view; numerical indicators view; customizable theme; preset notes; and custom notes.

Testing the Cloud and Mobile-Based App

The mobile and cloud-based systems was then piloted in a usability evaluation by health care providers and people with T2DM. They were asked to use the app for 10 days and complete the UEQ questionnaire and give their feedback and suggestions.

In total, 14 patients (Table 1) and 7 health care providers took part in the study. The age range of patients was 24 to 53 years, most of them were female (10/14, 71%). The age range of health care providers was 35 to 42 years. Most of the health care providers who took part in this study were female (6/7, 85%).

Table 1. Results of the User Experience Questionnaire completed by patients (n=14).

Item	Median (SD)	IQR (Q3–Q1)
Attractiveness	1.56 (0.43)	0.56
Perspicuity	2.35 (0.82)	1.08
Efficiency	1.76 (0.50)	0.66
Dependability	1.70 (0.46)	0.56
Stimulation	1.96 (0.48)	0.64
Novelty	2.30 (0.60)	0.80

The highest median belongs to perspicuity and novelty and the lowest for attractiveness. The reason for the high perspicuity measure seems to be due to the simple design of the system. In designing this system, the features and modules were put together simply and clearly and enough explanations were written about each module for users. The point with the next highest measure is novelty. As mentioned in the previous sections of the paper, the existence and use of mobile-based systems in patient care and monitoring in Iran are still in its infancy. Therefore, when patients have used this system on their phones, its novelty seemed like one of the most prominent features. This measure has led to an increase in patients’ stimulation to use it.

The lowest level is related to the measure of attractiveness. Perhaps one of the main reasons for this is related to the simple design of the system. The textual content of the system instead of its visual content was applied. In other words, patients used this system by looking at the numbers in the charts, receiving recommendations and text messages, and selecting and typing the name of foods and physical activities in text form instead of images. Also, the lack of use of sound and music in the relevant sections could have reduced the attractiveness for patients. This content can increase the system attractiveness to users.

The UEQ offers such a benchmark, which contains the data of previous product evaluations with these results. The benchmark classifies a product into 5 categories for each measure. Figure 6 illustrates this comparison for mobile-based system.

A total of 4 nurses and 3 physicians collaborated to evaluate the cloud-based app. Table 2 shows the median, interquartile range, and standard deviation of each measure asked among the health care providers. The highest average is for perspicuity and the lowest for attractiveness (Table 2).

Figure 6. Comparison chart of average mobile app system measurements with previous studies.

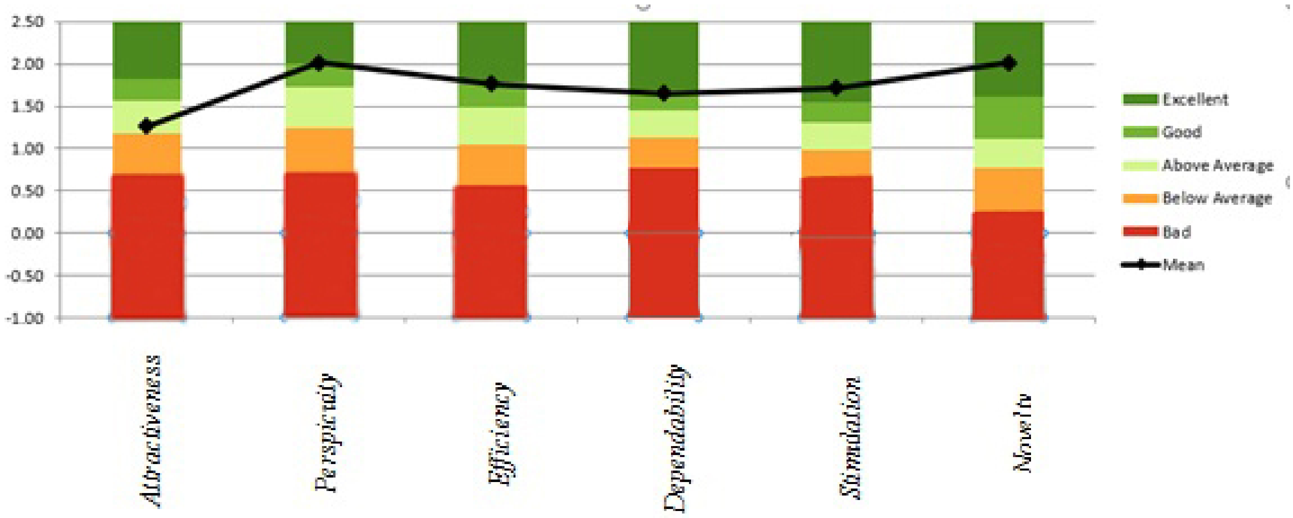


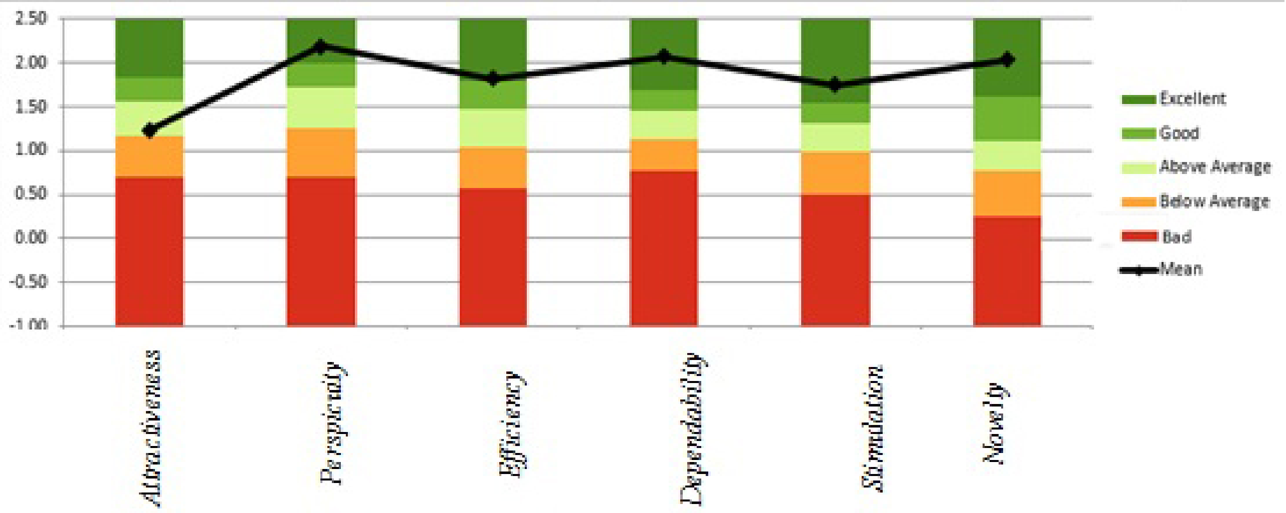
Table 2. Results of the User Experience Questionnaire completed by health care providers (n=7).

Item	Median (SD)	IQR (Q3–Q1)
Attractiveness	1.53 (0.25)	0.32
Perspicuity	2.45 (0.62)	0.62
Efficiency	2.05 (0.34)	0.44
Dependability	2.15 (0.45)	0.62
Stimulation	1.95 (0.43)	0.60
Novelty	2.05 (0.30)	0.40

The highest scores were given for perspicuity. The reason for this is that we intended to design the system as clearly as possible for health care providers. We aimed to make it user-friendly so that working with it was not time-consuming. In addition, they were provided with sufficient explanations for each module. One of the reasons for the high dependability measure for health care providers could be the ability to view and access the data online and instantly. They could check the lists of educational recommendations and messages in addition to monitoring data and the process of changing stages of patients in the form of charts. In the usual state of care, it was not always been possible to observe the retrospective data in this way or it may have taken more time.

Similar to the results of the usability evaluation for patients, the lowest level of measurement is related to attractiveness. Perhaps one of the reasons for this is the simple design of the system. As mentioned, the appearance of the system was as simple as possible, so the use of images and a lot of colors was avoided. Figure 7 illustrates this comparison for the web-based system.

Figure 7. Comparison chart of average web-based system measurements with previous studies.



Discussion

Principal Findings

Using a cloud- and mobile-based system had many advantages for people with diabetes and health care providers. These apps could be helpful for persons receiving online education in self-management, relevant messages, and tips for their needs. Health care providers could also get an overview to conveniently obtain information about a person's health and behavior status. Since cloud-based systems and mobile apps can be used as a tool for remotely monitoring and managing a person's health, this feature can be effective in providing organized care for many other chronic diseases [1].

The app that we developed in this study differs from previous Iranian apps in a number of key aspects. It was developed based on a behavioral framework and provided customized messages and recommendations. Using this app, people with T2DM and their health care providers had multifunctional apps that enabled them to enter data in the logbook, view results in charts and tables, view them in more extensive time trends, receive generalized and customized messages and education, receive reminders and alarms, and have a concise useful clinical record. Apps that have multiple functions appear to be more likely to be used than those that have only a single function. The most common functions in diabetes apps include data documentation, data transfer, information collection, analysis, and reminders [31]. However, there must be a balance between the proper performance of an app and its user digital literacy level. The number and type of functions of the proposed app in this study were based on the applicable standards in the field and opinion of the specialists that we consulted.

Several studies have proven the effectiveness and benefits of charting data for diabetes care [32-35]. One of the features used in our proposed app was the use of charts to represent the data. Using visual approaches such as charts for blood sugar, physical activity, and calorie intake can be valuable to both people with diabetes and health care providers because they will be able to understand the information hidden in the data at a glance and make informed decisions better and faster by observing the process of change. Well-designed features in the mobile-based systems such as customized education demonstrated the potential for enhancing a person's self-management outcomes. Developing such an app with customized feedback may help increase a person's participation in telemedicine interventions as it provides relevant, timely, and specific feedback. Individual and behavioral characteristics of individuals are unique to them and can affect the level of self-management of each person differently. Therefore, tailored interventions can be more appropriate for each person's individual needs [36].

In the study by Broderick and Haque [37], it was shown that there are 3 main barriers to mHealth interventions. These barriers include technical and human resources in organizations to support the implementation of these projects, lack of sufficient funding to finance investment in mobile technology solutions, and the challenges of merging health solutions with electronic health records and others. These barriers can also be compounded by user disinclination to cooperate and the fear of using tools, more common in developing countries. This can be attributed to a large number of specialist visits and the lack of sufficient time. People with diabetes themselves also acknowledge the concerns of using these apps, which may be related to their experiences of using unreliable online tools. This is a reasonable concern because few apps are research-based and even may be developed without any guideline-based or expert-approved content [22]. The results of this study showed that users were generally satisfied with the app. This overall positive feedback can be effective and promising for the first steps of implementing such apps in Iran. However, training on this issue and providing the right context can always be a challenge. Therefore, there should be more extensive research to respond to the growing needs under such circumstances. After this pilot study, further research is needed to investigate user attitudes and the likelihood of use in a larger population.

Since the researchers aimed to design and develop a new app rather than investigate the effects of using a predesigned system, the small number of users was not an obstacle to this goal. However, this number was really limited. Only 21 users (include

14 people with diabetes and 7 health care providers) used and evaluated this cloud- and mobile-based system, and the results cannot be generalized to all people diagnosed with T2DM. In other words, the purpose of this study was to design an mHealth intervention by using TTM, and the recruitment of potential users was required only for the initial deployment of the program. We performed only a usability evaluation in the real environment. The clinical outcomes of this intervention have not been studied yet and will be the topic of our next studies. Further clinical trial studies are needed to reveal the efficacy of this product compared with routine care.

Conclusion

This study describes the development of a cloud-based and mobile-based system for people with diabetes and their health care providers. We used TTM as the theoretical foundation of this system and tailored massaging for improving the acceptability of the system by the users. Although some positive evaluation metrics were observed, a limited sample size did not allow for any concrete conclusions to be drawn from this study's findings. More in-depth exploratory analysis of usability issues is needed to inform the design of clinical trials in this field.

Acknowledgments

This study was conducted as part of the PhD study of RS at Tehran University of Medical Science, Tehran, Iran. FF was financially supported by the Queensland Government through an Advance Queensland Fellowship.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Self-created questionnaire based on the transtheoretical model.
[DOCX File , 42 KB - jmir_v23i6e18167_app1.docx]

References

1. Ayatollahi H, Hasannezhad M, Fard H, Haghighi M. Type 1 diabetes self-management: developing a web-based telemedicine application. Health Inf Manag 2016 Apr;45(1):16-26. [doi: 10.1177/1833358316639456] [Medline: 28691565]
2. Mokhtari Z, Gheshlagh R, Kurdi A. Health-related quality of life in Iranian patients with type 2 diabetes: an updated meta-analysis. Diabetes Metab Syndr 2019;13(1):402-407. [doi: 10.1016/j.dsx.2018.10.007] [Medline: 30641733]
3. Cho N, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, et al. IDF Diabetes Atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018 Apr;138:271-281. [doi: 10.1016/j.diabres.2018.02.023] [Medline: 29496507]
4. Diabetes prevalence in the Middle East and North Africa region (estimates for 2017 and 2045). International Diabetes Federation. 2017. URL: https://www.idf.org/index.php?option=com_attachments&task=download&id=1509:Atlas-8e-poster-MENA-region [accessed 2019-10-30]
5. Hadipour M, Abou Alhasani F, Molavi Vardanjani H. Health-related quality of life in patients with of type II diabetes in Iran. Payesh 2013;12(2):135-141.
6. International Diabetes Federation Middle East and North Africa region members. URL: https://idf.org/our-network/regions-members/middle-east-and-north-africa/members/35-iran.html [accessed 2019-10-30]
7. Esteghamati A, Larijani B, Aghajani M, Ghaemi F, Kermanchi J, Shahrami A, et al. Diabetes in Iran: prospective analysis from First Nationwide Diabetes Report of National Program for Prevention and Control of Diabetes (NPPCD-2016). Sci Rep 2017 Oct 18;7(1):13461 [FREE Full text] [doi: 10.1038/s41598-017-13379-z] [Medline: 29044139]
8. Chatterjee S, Davies M, Heller S, Speight J, Snoek F, Khunti K. Diabetes structured self-management education programmes: a narrative review and current innovations. Lancet Diabetes Endocrinol 2018 Feb;6(2):130-142. [doi: 10.1016/s2213-8587(17)30239-5]
9. Bellei EA, Biduski D, Lisboa HRK, De Marchi ACB. Development and assessment of a mobile health application for monitoring the linkage among treatment factors of type 1 diabetes mellitus. Telemed J E Health 2020 Feb;26(2):205-217. [doi: 10.1089/tmj.2018.0329] [Medline: 30724717]
10. Goyal S, Lewis G, Yu C, Rotondi M, Seto E, Cafazzo JA. Evaluation of a behavioral mobile phone app intervention for the self-management of type 2 diabetes: randomized controlled trial protocol. JMIR Res Protoc 2016 Aug 19;5(3):e174 [FREE Full text] [doi: 10.2196/resprot.5959] [Medline: 27542325]
11. Weinstein R, Lopez A, Joseph B, Erps K, Holcomb M, Barker G, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
12. Hebden L, Cook A, van der Ploeg HP, Allman-Farinelli M. Development of smartphone applications for nutrition and physical activity behavior change. JMIR Res Protoc 2012 Aug 22;1(2):e9 [FREE Full text] [doi: 10.2196/resprot.2205] [Medline: 23611892]

13. Hood M, Wilson R, Corsica J, Bradley L, Chirinos D, Vivo A. What do we know about mobile applications for diabetes self-management? A review of reviews. J Behav Med 2016 Dec;39(6):981-994. [doi: 10.1007/s10865-016-9765-3] [Medline: 27412774]
14. Sawesi S, Rashrash M, Phalakornkule K, Carpenter J, Jones J. The impact of information technology on patient engagement and health behavior change: a systematic review of the literature. JMIR Med Inform 2016 Jan 21;4(1):e1 [FREE Full text] [doi: 10.2196/medinform.4514] [Medline: 26795082]
15. Hood KK, Hilliard M, Piatt G, Ievers-Landis CE. Effective strategies for encouraging behavior change in people with diabetes. Diabetes Manag 2015 Nov;5(6):499-510. [doi: 10.2217/dmt.15.43]
16. Pal K, Eastwood S, Michie S, Farmer A, Barnard M, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database of Systematic Reviews 2013:1. [doi: 10.1002/14651858.cd008776]
17. Prochaska J, DiClemente C. Transtheoretical therapy: toward a more integrative model of change. Psychother Theory Res Pract 1982;19(3):276-288. [doi: 10.1037/h0088437]
18. Jackson R, Asimakopoulou K, Scammell A. Assessment of the transtheoretical model as used by dietitians in promoting physical activity in people with type 2 diabetes. J Hum Nutr Diet 2007 Feb;20(1):27-36. [doi: 10.1111/j.1365-277x.2007.00746.x]
19. Rangraz Jeddi F, Nabovati E, Hamidi R, Sharif R. Mobile phone usage in patients with type II diabetes and their intention to use it for self-management: a cross-sectional study in Iran. BMC Med Inform Decis Mak 2020 Feb 07;20(1):24 [FREE Full text] [doi: 10.1186/s12911-020-1038-y] [Medline: 32033560]
20. Zamanzadeh V, Zirak M, Hemmati Maslakpak M, Parizad N. Distance education and diabetes empowerment: a single-blind randomized control trial. Diabetes Metab Syndr 2017 Nov;11 Suppl 1:S247-S251. [doi: 10.1016/j.dsx.2016.12.039] [Medline: 28034693]
21. Ross J, Gao J. Overcoming the language barrier in mobile user interface design: a case study on a mobile health app. arXiv. Preprint posted online on May 16, 2016. URL: https://arxiv.org/pdf/1605.04693 [accessed 2021-05-21]
22. Salari R, Niakan Kalhori SR, Fatehi F, Ghazisaeedi M, Nazari M. Determining minimum set of features for diabetes mobile apps. J Diabetes Metab Disord 2019 Dec;18(2):333-340 [FREE Full text] [doi: 10.1007/s40200-019-00417-y] [Medline: 31890658]
23. Peeples M, Tomky D, Mulcahy K, Peyrot M, Siminerio L, AADE Outcomes ProjectAADE/UMPC Diabetes Education Outcomes Project. Evolution of the American Association of Diabetes Educators' diabetes education outcomes project. Diabetes Educ 2007;33(5):794-817. [doi: 10.1177/0145721707307615] [Medline: 17925584]
24. Highstein GR, O'Toole ML, Shetty G, Brownson CA, Fisher EB. Use of the transtheoretical model to enhance resources and supports for diabetes self management: lessons from the Robert Wood Johnson Foundation Diabetes Initiative. Diabetes Educ 2007 Jun;33 Suppl 6:193S-200S. [doi: 10.1177/0145721707304476] [Medline: 17620401]
25. Parham M, Riahin A, Jandaghi M, Darivandpour A. Self care behaviors of diabetic patients in Qom. Qom University of Medical Science Journal 2016 Apr 05:81-87.
26. Kreuter M, Strecher V, Glassman B. One size does not fit all: the case for tailoring print materials. Ann Behav Med 1999 Dec;21(4):276-283. [doi: 10.1007/bf02895958]
27. Delavari A, Mahdavi A, Norouzinejad A. Country Disease Prevention and Control Program. Minstry of Health and Medical Education (MoHME). 2005. URL: https://fhc.sums.ac.ir/files/ghvagir/ayin%20name/diabetes_guideline_91_final_-_Copy.pdf [accessed 2019-10-30]
28. Resources for people living with diabetes. Association of Diabetes Care & Education Specialists. URL: https://www.diabeteseducator.org/living-with-diabetes/aade7-self-care-behaviors [accessed 2019-10-27]
29. Diabetes education. International Federation of Diabetes. URL: https://www.idf.org/our-activities/education.html [accessed 2019-10-27]
30. Diabetes type 2: living with diabetes. American Diabetes Association. URL: http://www.diabetes.org/diabetes-basics/type-2/?loc=util-header_type2 [accessed 2019-10-27]
31. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr 09;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
32. Chen Y, Kao S, Tan M, Ning Y, Salloway M, Wee H, et al. Feasibility of representing adherence to blood glucose monitoring through visualizations: a pilot survey study among healthcare workers. Int J Med Inform 2018 Dec;120:172-178. [doi: 10.1016/j.ijmedinf.2018.09.006] [Medline: 30409342]
33. Tao D, Yuan J, Qu X. Presenting self-monitoring test results for consumers: the effects of graphical formats and age. J Am Med Inform Assoc 2018 Aug 01;25(8):1036-1046 [FREE Full text] [doi: 10.1093/jamia/ocy046] [Medline: 29762686]
34. Feller D, Burgermaster M, Levine M, Smaldone A, Davidson P, Albers D, et al. A visual analytics approach for pattern-recognition in patient-generated data. J Am Med Inform Assoc 2018 Oct 01;25(10):1366-1374 [FREE Full text] [doi: 10.1093/jamia/ocy054] [Medline: 29905826]
35. Zhang Y, Chanana K, Dunne C. IDMVis: Temporal Event Sequence Visualization for type 1 diabetes treatment decision support. IEEE Trans Visual Comput Graphics 2019 Jan;25(1):512-522. [doi: 10.1109/tvcg.2018.2865076]

36. Cassimatis M, Kavanagh D, Smith A. Perceived needs for supported self - management of type 2 diabetes: a qualitative investigation of the potential for a web - based intervention. Aus Psychol 2020 Nov 12;49(2):75-85. [doi: 10.1111/ap.12050]
37. Broderick A, Haque F. Mobile health and patient engagement in the safety net: a survey of community health centers and clinics. Issue Brief (Commonw Fund) 2015 May;9:1-9. [Medline: 26040018]

Abbreviations

BG: blood glucose
mHealth: mobile health
TTM: transtheoretical model
T2DM: type 2 diabetes mellitus
UEQ: User Experience Questionnaire

Edited by G Eysenbach; submitted 08.02.20; peer-reviewed by R Rabiei, S Sankaran, JJ Mira, M Nomali, L Wang, A Sudaryanto; comments to author 20.04.20; revised version received 29.05.20; accepted 07.07.20; published 02.06.21.

Please cite as:
Salari R, R Niakan Kalhori S, GhaziSaeedi M, Jeddi M, Nazari M, Fatehi F
Mobile-Based and Cloud-Based System for Self-management of People With Type 2 Diabetes: Development and Usability Evaluation
J Med Internet Res 2021;23(6):e18167
URL: https://www.jmir.org/2021/6/e18167
doi:10.2196/18167
PMID:34076579

Original Paper

Feasibility and Acceptability of a Mobile Phone App Intervention for Coping With Cancer as a Young Adult: Pilot Trial and Thematic Analysis

Hanneke Poort[1], PhD; Annelise Ryan[1], MPH; Katelyn MacDougall[1], LICSW; Paige Malinowski[1], BA; Anna MacDonald[1], BA; Zach Markin[2], BSE; William Pirl[1], MPH, MD; Joseph Greer[3], PhD; Karen Fasciano[1], PsyD

[1]Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Boston, MA, United States

[2]HTD Health, New York, NY, United States

[3]Center for Psychiatric Oncology and Behavioral Sciences, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Hanneke Poort, PhD
Department of Psychosocial Oncology and Palliative Care
Dana-Farber Cancer Institute
450 Brookline Avenue
Boston, MA, 02215
United States
Phone: 1 6175827617
Email: hanneke_poort@dfci.harvard.edu

Abstract

Background: Many young adult patients do not receive adequate psychosocial services to help them cope with cancer.

Objective: This study aims to assess the feasibility and acceptability of a smartphone app (*iaya*) intervention that was designed to create an engaged community of young adult patients and help them learn emotional coping skills.

Methods: For this single-group pilot trial, 25 young adult patients aged 18-39 years who were receiving active cancer treatment were asked to use the *iaya* app for 12 weeks. To collect app use data, we used Mixpanel, an analytics platform for apps. Feasibility was assessed through rates of app sessions and the number of coping exercises engaged, and intervention acceptability was evaluated by using an app usability questionnaire and through qualitative interviews at study completion. We collected patient-reported outcome data at baseline and at week 12 to explore self-efficacy for coping with cancer, self-efficacy for managing emotions, perceived emotional support, and quality of life.

Results: Baseline patient-reported outcome data indicated that participants scored relatively low on perceived emotional support but reasonably high on self-efficacy for coping with cancer and managing emotions as well as quality of life. Participants had a mean of 13 app sessions (SD 14) and 2 coping exercises (SD 3.83) in 12 weeks. Only 9% (2/23) of participants met our combined feasibility definition of ≥10 app sessions and ≥3 coping skills from different categories. The participants' mean usability score was 73.7% (SD 10.84), which exceeded our predefined threshold of ≥70%, and qualitative feedback was generally positive.

Conclusions: Although perceived acceptable by patients, the *iaya* smartphone app did not meet the a priori feasibility criteria as a stand-alone app intervention. Future studies should screen participants for unmet coping needs and consider integrating the app as part of psychosocial care for young adult patients.

(J Med Internet Res 2021;23(6):e25069) doi:10.2196/25069

KEYWORDS

mobile phone; mobile phone application; cancer; feasibility

Introduction

Background

Young adults (aged 18-39 years) experience a unique set of psychosocial needs while receiving treatment for cancer. Young adulthood, a phase of life critical to development, is characterized by establishing identities, negotiating independence from parents, completing education, starting a career, and making life decisions about relationships and family [1]. A cancer diagnosis can interrupt these milestones and severely alter a young adult's life and affect their psychological well-being [2,3]. Moreover, young adults are at a unique stage in their emotional, cognitive, and social development, and many young adults with cancer report high levels of cancer-related distress and other psychological issues, such as isolation, anxiety, and depression [4].

The National Comprehensive Cancer Network guidelines call for developmentally appropriate psychosocial support services for young adults with cancer. These services must be flexible in approach; provide age-appropriate information on important topics such as financial health, fertility, work, body image, and sex; and recognize the importance of peer support [5]. Despite the existing body of research in this field and the subsequent recommendations, a substantial proportion of young adults (41%) report unmet needs for psychosocial support during the first year following diagnosis [2]. Failure to meet such psychosocial support needs is associated with increased distress in this population [6]; therefore, providing these services is of utmost importance in the care of young adults with cancer.

Several factors may interfere with young adults' access to or utilization of psychosocial care services. One barrier is the lack of awareness among the general adult cancer population about the available services. In 2010, the National Health Interview Survey reported that 90% of patients did not know that psychosocial services existed or that they were available to them [7]. Furthermore, young adults can be difficult to engage in psychosocial services because of the stigma of accessing mental health services [8]. In addition, young adults typically have lives with multiple demands outside of coping with cancer, which may reduce a young adult's ability and willingness to attend on-site peer support activities and psychosocial care visits [9]. More specifically, young adults may be unable or less willing to participate in in-person activities that address their psychosocial needs because of ill-health, treatment side effects, and the considerable effort required to manage and complete treatment [10,11]. Therefore, it is important to explore interventions that promote psychosocial support services that young adults can easily access and incorporate into their lives without jeopardizing competing priorities.

Solutions involving technology and social media show promise for delivering age-appropriate psychosocial care to young adults with cancer, as they can also eliminate some of the access barriers. As of 2019, 96% of young adults aged 18-29 years reported having a smartphone, and 100% of millennials aged 23-38 years said that they used the internet [12,13]. In addition, the majority of young adults are comfortable using social media platforms, and over half of all 18-29–year-olds who are on the web access these types of sites daily [14]. Specifically, within the cancer sphere, young adults have expressed interest in smartphone apps and social media interventions aimed at providing mental health care and peer support [15,16]. Other investigators have also recognized the potential of delivering cancer-related care via smartphone app. However, most cancer-related apps aim to raise awareness and provide information, whereas a smaller subset targets prevention, early detection, management of cancer, and social networking [17,18]. Very few studies on cancer-related apps have attempted to explore the effectiveness and utility of delivering psychosocial care, and even fewer studies have explored the delivery of age-appropriate psychosocial care for young adults with cancer [19,20].

Objective

In this work, we describe the pilot testing of a smartphone app intervention (*iaya*) in young adults who were receiving active cancer treatment. The *iaya* app was designed with input and feedback from young adults to build an engaged community, to develop coping skills, and to encourage personal development. The goal of this study is to assess the feasibility (defined as number of app sessions and coping skills engaged) and acceptability of the *iaya* app intervention among young adults with cancer.

Methods

Study Setting and Participants

We conducted a single-arm pilot study in the lymphoma, leukemia, breast, melanoma, sarcoma, gastrointestinal, and neuro-oncology clinics at the Dana-Farber Cancer Institute in Boston, Massachusetts, between November 11, 2019, and March 13, 2020. The eligibility criteria were as follows: had a diagnosis of cancer, were receiving active cancer treatment, were under the care of a Dana-Farber oncologist, aged 19-39 years, had access to a smartphone (iOS or Android), and had a willingness to use the *iaya* app and complete study surveys as well as a qualitative exit interview. Exclusion criteria included the following: an inability to provide informed consent in English and cognitive or neurological impairments that might preclude study participation, as evaluated by the research study staff or oncology provider. We aimed to enroll 30 participants for this pilot feasibility trial over a 6-month enrollment period but suspended recruitment 2 months earlier because of halted research operations related to the COVID-19 pandemic.

Enrollment

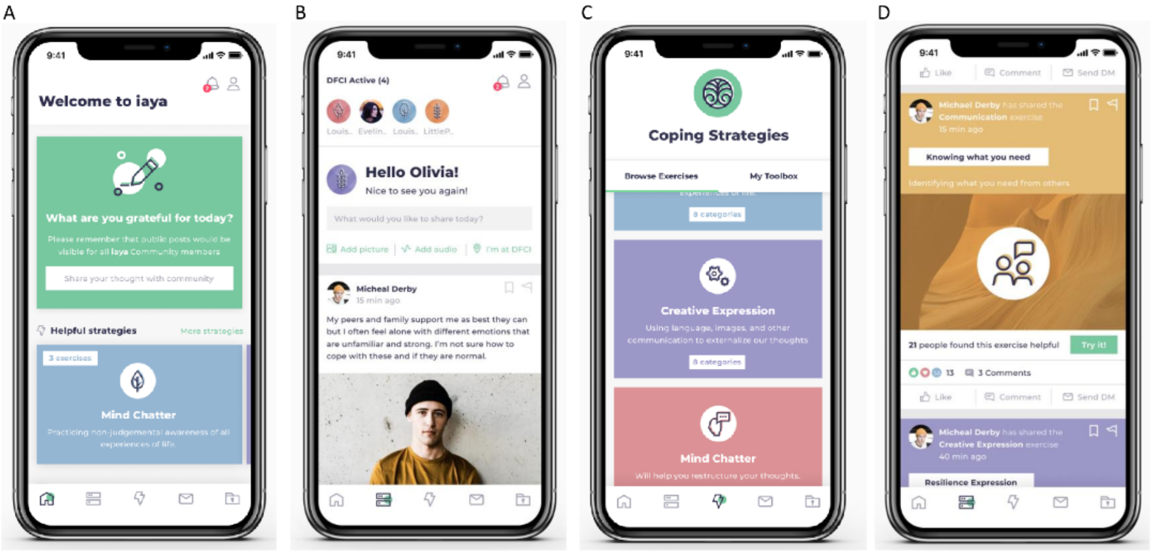
We conducted a limited review of clinic schedules to identify potentially eligible patients. A research staff member (AR) contacted the patient's oncology provider to confirm eligibility and request permission to approach the patient at an upcoming clinic visit. Patients approved for approach for enrollment were invited by a research staff member to enroll in the study. Patients were informed during the consent process that the app was freely available outside of study participation, as the app was first launched as a clinical tool on April 6, 2019, as part of the Dana-Farber Young Adult Program. Participants were specifically informed about the data that would be collected from their phone, the methods used to secure and encrypt these data, and the information that would be used for the study. All participants provided written informed consent for all the study procedures. The participants did not receive financial incentives for completion of study procedures. The study was approved by the Institutional Review Board of the Dana-Farber/Harvard Cancer Center.

Intervention

The *iaya* intervention was a smartphone app intervention designed to improve the psychological care for young adults with cancer, and it consisted of several components. Clinical content for the *iaya* app was developed by mental health clinicians specializing in the care for young adult patients with cancer (KM and KF). We sought to accomplish our goal of improving young adult–specific psychological care by providing users with an app that included a combination of psychoeducational resources, coping skills training, and the opportunity to connect and share personal content with peers. All these components have theoretical underpinnings, including the sense of coherence theory [21,22], coping theory [23], and social support theory [24,25]. In addition, the opportunity to connect and share personal content with peers was also endorsed as a cancer-specific need by our app development stakeholder group (refer to the *App Development* section). The *iaya* app also included a geolocation feature that users could opt for if they were at Dana-Farber and which was meant to serve as a *virtual waiting room* to connect with other young adults with cancer on the app and at Dana-Farber at the time.

Figure 1 displays the interface and 5 main features of the app: the home page, community feed, private messaging, private feed, and coping exercises section. The dynamic *home page* had a rotating community question that users could respond to; a shortcut to the coping exercises; and the YAP Daily Post, which featured a different young adult resource every day. Notifications could be enabled or disabled based on an individual user's preference. If left enabled, users would receive a notification every time the community question rotated on the home page, which was 2-3 times a week. Similarly, users could either create a profile with their real name or create a username based on their individual preference. In the *community feed*, users could post information on a public forum and comment or react to other users' posts. If users wanted to connect with other users privately, they were able to send direct messages, in the *private messaging* section, to individuals that they met on the community feed. The *private feed* served as a space for users to note their thoughts and save meaningful information they found on the community feed. The *coping exercises* section in the app featured exercises from evidence-based therapies used in clinical practice to help treat people with anxiety and mood difficulties (ie, cognitive behavioral therapy, dialectical behavior therapy, acceptance and commitment therapy, and mindfulness-based stress reduction) as well as strategies that help facilitate emotional resilience.

Figure 1. Features and interface of the iaya app. Main app features displayed on a (fictive) user's smartphone, as follows: iPhone smartphone screens displaying the home page with a community question (A) and the community feed with active users on the web (B) and the option to check into the virtual waiting room through the geolocation feature "I'm at DFCI." iPhone smartphone screens displaying the community feed with a user's shared output from a coping exercise (C) to which users can respond by liking, commenting, direct messaging, or trying it themselves and the coping strategies section (D) in which users have the option to browse exercises or go to "My Toolbox," where they shortlisted previously completed exercises for easy access. DFCI: Dana-Farber Cancer Institute.



My peers and family support me as best they can but I often feel alone with different emotions that are unfamiliar and strong. I'm not sure how to cope with these and if they are normal.

In the coping exercises section, users had access to 131 unique exercises that were divided into 6 categories: Values (identifying things we care most about), Communication (clear assertive communication to optimize self-advocacy and support from others), Let It Be (building skills to limit struggles with aspects of life we cannot control), Mind Chatter (help restructure your thoughts), Take Action (feel fully connected to your body, especially when you feel triggered by anxiety or worry), and Expression (using language, images, and other communication to externalize and express thoughts). Textbox 1 provides a more detailed description of these categories and their content. Coping strategies included both educational information and active exercises to practice the strategy. The active exercises allowed the user to generate an output to share on the community feed.

Textbox 1. Coping exercise categories and content.

Values: "Identifying the Things We Care Most About"

- Exercises to help the user concentrate on ways to define and focus on what is important to them and how this can help with enhancing positive emotions and shift away from negative emotions. Users will explore the relationship between values and goals. Finally, users are asked to strengthen their identity outside of their illness by focusing on values and redefining goals to meet values.

Communication: "Clear Assertive Communication to Optimize Self-advocacy and Support From Others"

- Exercises designed to help young people enhance positive strategies for adjusting to needing help and defining and asking for what they need. This section also includes exercises for skills in setting limits and boundaries for self-care as well as improving open communication to enhance social support. In addition, users are asked to practice communication that conveys thanks and gratitude. Finally, users are coached through strategies for disclosure about cancer.

Let It Be: "Building Skills to Limit Struggles With Aspects of Life We Can't Control"

- Exercises aimed at identifying and normalizing emotions, expression of emotions, and defining what can and cannot be controlled. Users learn more about acceptance and accepting emotions as well as being intentional with emotions.

Mind Chatter: "Will Help You Restructure Your Thoughts"

- Exercises aimed at restructuring thoughts that may be cognitive errors. Common cognitive errors are highlighted and include thinking filter, blaming, jumping to conclusions, control myths, emotional reasoning, and personalization. For each of these, users are asked if they can relate to the error, and if so, the user is asked to review examples, restructure a thought that is presented to them, and come up with their own example and a restructured solution.

Take Action: "Feel Fully Connected to Your Body, Especially When Triggered by Anxiety or Worry"

- Exercises include behavioral strategies that reduce the stress response and enhance mindfulness. Some of these strategies promote present moment awareness, whereas other guided imagery strategies take the user into a created visual image that enhances relaxation. Basic breathing is also addressed in this category, as are progressive muscle relaxation, grounding, and mindful action techniques.

Expression: "Using Language, Images, and Other Communication to Externalize Our Thoughts"

- Exercises in this category provide the user with prompts for posting written, visual, or auditory contributions to express themselves around young adult's identity development in general or to the challenges and impact of cancer. Users are provided with prompts to promote expression of resilience-based principles such as gratitude, defining successes, meaningful actions, and priorities.

App Development

We developed the *iaya* app as a stand-alone intervention with input from mental health clinicians of the Young Adult Program at Dana-Farber (KF and KM), other psychosocial clinicians at Dana-Farber, and engineers from HTD Health (New York, New York), with funding from the Oak Foundation. In addition, we had 8 stakeholder meetings with young adult patients between November 2016 and March 2019. Three early meetings (2016-2017) were focused on soliciting general thoughts about and preferences for technology solutions for providing psychosocial care to young adults. An additional 5 meetings (2018-2019) were specifically focused on the user testing of prototypes and beta versions of the app. After the launch of the app in April 2019 (ie, 6 months before opening the feasibility study), we invited 9 young adult patients from the stakeholder meetings to remain involved as *super users*. We encouraged these *super users* to continue to beta test and provide real-time user feedback through an in-app feedback and bug reporting tool for smartphone apps (Instabug). In addition, we shared suggestions with them via weekly emails for 2 months on how to increase user engagement on the app, such as responding to other users' contributions and posting content.

App Use

The *iaya* app is a microsocial network that exposes a native smartphone app for iOS and Android to young adult patients with cancer and includes a separate web interface for community and content management. All data are encrypted while in transit and stored on an encrypted database server provided by the Partners Health System. All data were collected and stored in a Health Insurance Portability and Accountability Act–compliant manner. We used the Mixpanel analytics platform to collect data on participants' use behavior. We gathered information unique to users, yet fully deidentified, to understand the frequency of app use. We also recorded the number of coping exercises engaged. Deidentified use data were sent directly from a user's mobile device to Mixpanel, where it was stored with a unique and anonymous identifier for later analyses.

Procedures and Reminders

Participants were asked to download the freely available *iaya* app and recommended to use the app ≥3 times per week. During the first 3 weeks of study participation, a research assistant with a master's degree in public health (AR) conducted weekly phone calls to see if participants had any technical issues with accessing or using the app. Thereafter, the participants received one more phone call at 6 weeks. If phone calls were unanswered, a voicemail message was left with the instruction to return the call, in case of any technical difficulties with the app. In addition to technical support, these phone calls also served as reminders of the *iaya* app. This hybrid approach was adopted in an attempt to minimize early intervention dropouts and to overcome poor uptake of the smartphone app.

Revisions and Updating

No major changes to features or content were made to the app during the research study. At the time of study commencement (November 11, 2019), participants downloaded the 1.0.20 version of the *iaya* app. Early in the study, an app update was required to ensure correct data collection by Mixpanel. We released version 1.0.21 on December 11, 2019, at which point the research assistant contacted the 12 participants actively enrolled in the study to update their app to the most recent version. All but one participant updated their app to the new version, which resulted in a loss of use data from this participant. Three months into the study, we required another update because data collection to measure the number of app sessions through Mixpanel was incompatible with the Android platform. We released version 1.0.22 on March 16, 2020, to correct this unforeseen issue, which was downloaded by 5 participants. Finally, version 1.0.23 was released 1 week later, with minor enhancements to improve the *currently online* feature of the app, and this version was downloaded by 3 participants during the study period.

Measures

Demographics

Participants completed a basic sociodemographic questionnaire at enrollment (eg, age, marital status, children, race, education, employment, and religion). Clinical information, including primary disease site, time since diagnosis, and current cancer treatment, was extracted from the electronic health record.

Feasibility

Feasibility was defined as ≥75% of participants with ≥10 app sessions (ie, a user opens the app and has it open for at least 10 seconds before the app is closed or moved to the background) and ≥3 coping skills opened from different categories over a period of 12 weeks. We corrected the auto-populated Mixpanel app session data to eliminate multiple app sessions within a 5-minute window.

Acceptability

Acceptability was defined as ≥70% mean app usability score derived from the app usability questionnaire that participants filled out upon study completion. This questionnaire was adapted from the 10-item System Usability Scale [26] to a simplified 6-item scale. Items are scored on a 5-point Likert scale (1=strongly disagree; 5=strongly agree), and each item's score contribution ranges from 0 to 4. Sum scores were converted to a 0-100 range; higher scores indicated a higher perception of usability. Scores ≥70% can be considered to have good usability. In addition, patients participated in qualitative exit interviews with a member of the study team (AR) to elicit feedback on their experiences with the app intervention. The interviews were audio-recorded and transcribed.

Patient-Reported Outcomes

We included 4 measures to explore self-efficacy for coping with cancer, self-efficacy for managing emotions, perceived emotional support, and quality of life at baseline and 12 weeks. The study data were collected outside of the smartphone app and were managed using REDCap (Research Electronic Data Capture) hosted at Partners Healthcare.

The Cancer Behavior Inventory-Brief Version (CBI-B) is a 12-item questionnaire that assesses self-efficacy regarding different coping skills in the context of cancer (eg, maintaining a positive attitude, asking physicians questions, seeking consolation, and coping with physical changes) [27,28]. Items are scored on a 1-9 Likert scale (1=not at all confident; 9=totally confident). We calculated a total summed score (range 12-108), with higher scores indicating greater self-efficacy for coping, which is associated with better adjustment to cancer.

We used the 8-item Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form v1.0-Self-Efficacy for Managing Emotions 8a to assess a participant's level of confidence in managing symptoms of anxiety, distress, discouragement, disappointment, and negative feelings [29,30]. Items are scored on a 1-5 Likert scale (1=I am not at all confident; 5=I am very confident). We also used the 8-item short form PROMIS Short Form v2.0-Emotional Support 8a to assess perceived feelings of being cared for and valued as a person [31]. Items are scored on a 1-5 Likert scale (1=never; 5=always). For both PROMIS measures, we calculated a summed score and converted the total raw score into a T-score for each participant [32]. A higher PROMIS T-score represents more of the concept being measured (ie, greater self-efficacy for managing emotions and more emotional support). The standardized T-score has a mean of 50 (SD 10), indicating the average for the general chronic condition population in United States.

The 27-item Functional Assessment of Cancer Therapy-General (FACT-G) was used to assess the general quality of life [33]. Items are scored on a 0-4 Likert scale (0=not at all; 4=very much). Subscale scores were summed to obtain a FACT-G total score (range 0-108), with higher scores indicating a better quality of life.

Data Analysis

Descriptive statistics were used to describe the characteristics of our sample and summarize patient-reported outcome data at baseline and at 12 weeks. In addition, we used an inductive (content-driven) approach to code and analyze the exit interviews using thematic analysis [34]. First, the study team members HP and AR performed open coding and memoing to evaluate the transcripts and repeatedly reviewed transcripts line by line to identify text related to participants' perceptions of the *iaya* app and suggestions for improvement. Next, these study team members (HP and AR) met regularly to systematically review and discuss coded contents to identify emergent themes and patterns and synthesize data across themes, both within and across participant types (eg, users vs nonusers of coping strategies).

Results

Overview

We identified 97 potentially eligible patients during the 4-month enrollment period (Figure 2). Of these, 5 were deemed inappropriate by their treating oncologist for the following reasons: *bad timing for the patient* (n=1), *too distressed* (n=1), and *patient disposition* (n=3). Among the 92 patients who were approached to participate, 25 (27%) provided consent and were enrolled in the study. The 3 most common reasons for declining study participation were not interested in participating in the study, not interested in study because the app was available otherwise, and the patient had too much going on at the time. As displayed in Table 1, the mean age was 28 years (SD 5); 14 participants were female, 14 were married, and 5 identified themselves as Black individuals or other people of color. A total of 23 participants had a college or advanced degree, and 13 participants were employed.

Figure 2. Flowchart of screening and enrollment. DFCI: Dana-Farber Cancer Institute; MD: medical doctor.

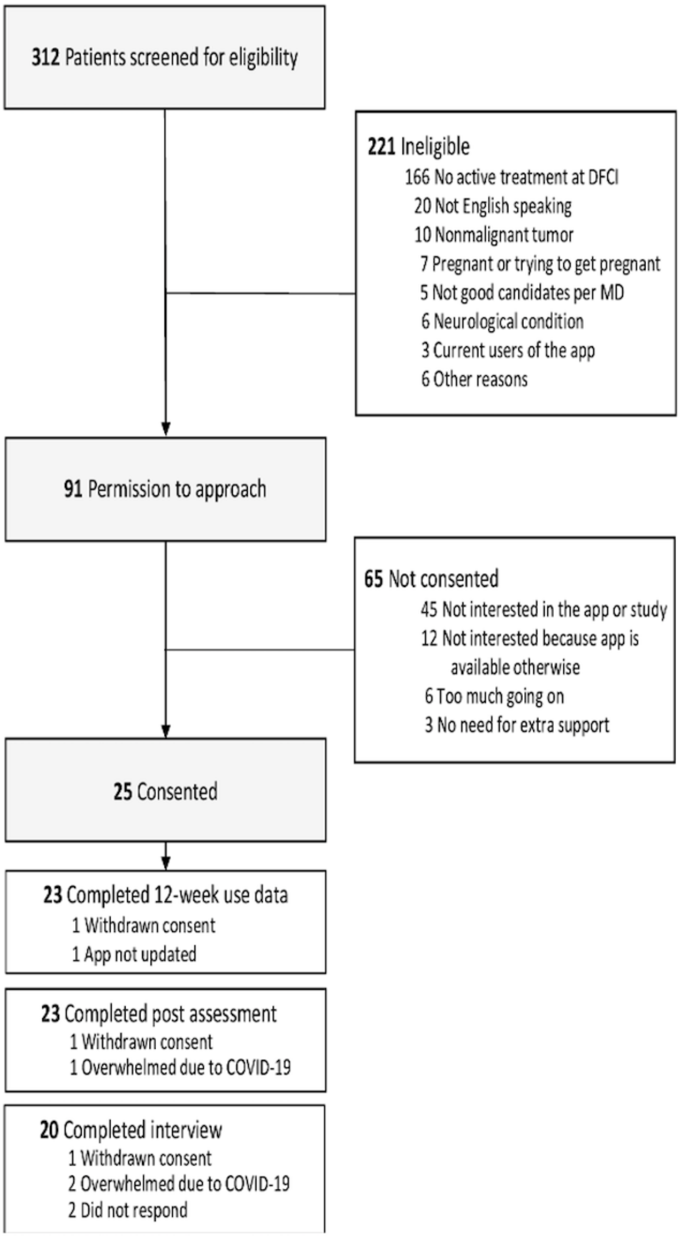


Table 1. Participant characteristics (N=25).

Characteristic	Value
Age (years), mean (SD)	28 (5)
Sex, n (%)	
Female	14 (56)
Male	11 (44)
Marital status, n (%)	
Married	14 (56)
Single, never married	7 (28)
Noncohabitating relationship	3 (12)
Divorced	1 (4)
Any children, n (%)	9 (36)
Race, n (%)	
White	20 (80)
Black	3 (12)
Other	2 (8)
Ethnicity, n (%)	
Non-Hispanic	23 (92)
Hispanic	1 (4)
Prefer not to answer	1 (4)
Education, n (%)	
High school	1 (4)
College student	2 (8)
College graduate or advanced degree	22 (88)
Employment, n (%)	
Employed	13 (52)
Disabled	8 (32)
Caring for home or family	2 (8)
Student	2 (8)
Primary disease site, n (%)	
Lymphoma	9 (36)
Breast	5 (20)
Sarcoma	5 (20)
Neuro	3 (12)
Gastrointestinal	1 (4)
Leukemia	1 (4)
Melanoma	1 (4)
Stage of disease, n (%)	
Early	11 (44)
Late	12 (48)
Not available[a]	2 (8)
Time since diagnosis (years), n (%)	
0	16 (64)
1-2	8 (32)

Characteristic	Value
≥3	1 (4)
Cancer treatment, n (%)	
Chemotherapy	22 (88)
Immunotherapy	2 (8)
Radiation	1 (4)

[a] Acute lymphoblastic leukemia or T-lymphoblastic lymphoma (n=1) and myelodysplastic or myeloproliferative neoplasm (n=1).

Feasibility

App use data were available for 23 participants. One participant had an older version of the smartphone app (1.0.20) that was incompatible with Mixpanel, and one participant completed only 3 weeks of the study period before withdrawing consent due to medical reasons. On average, participants had 13 app sessions (SD 14; range 1-50) in 12 weeks. A total of 48% (11/23) participants had ≥10 app sessions in 12 weeks. In addition, participants opened an average of 2 (SD 4; range 0-18) individual coping exercises. Two participants opened ≥3 coping skills from different categories. Thus, only 9% (2/23) of participants met our combined feasibility definition (ie, ≥10 app sessions and ≥3 coping skills from different categories over a period of 12 weeks).

Acceptability

Overview

Upon completion of the 12-week study period, 23 participants completed the app usability questionnaire. One participant did not complete any surveys, and one participant withdrew consent due to medical reasons. The mean app usability score from 23 participants was 73.7%, which exceeded our predefined acceptability threshold of ≥70%.

The interviews indicated that participants liked the app's overall design and features. For example, one patient said:

> *It looks pretty neat, and I can find what I'm looking for really fast, which is pretty good. I like it.* [P4]

Another participant noted:

> *It was pretty self-explanatory, easy to navigate through.* [P12]

The participants only had minor suggestions for technical or content improvements. Three themes emerged from the qualitative data and are described below. Additional illustrative quotes for each theme are provided in Textbox 2.

Textbox 2. Themes and illustrative quotes from exit interviews.

Social Support and Feelings of Isolation

- "I can definitely see people are trying to encourage each other. That is great. I also see several people after they finish treatment, they still come back and provided support which is super nice." (P4)
- "I think for me, the feeling of being alone in it was the biggest thing and then going into the app and seeing, I was just taken back at actually how many people my age are sick with it." (P7)
- "I'm kind of like a shy person, so I didn't really want to make the approach myself, but it was nice to see, I guess you perceive support from seeing other people going through similar struggles as you....I didn't make a connection with someone but if I wanted to meet them in person, I guess that would have been nice." (P11)
- "The solidarity was very important. I didn't necessarily post or message anybody but reading through what people wrote and knowing that other people are going through different things is very useful." (P15)
- "My peers at home are not going through this. Just knowing that other people are going through this and that I'm going to be okay. That I have people out there, that there are other people rooting for me....It made me feel less alone." (P19)

Use Versus Nonuse of Coping Exercises

- "I've mostly just used the coping mechanisms....A lot of the questions I felt calmed me down so that worked for me. They were pretty well done." (P9, user)
- "Favorite feature of the app, definitely. It helped with communicating with myself, so to speak, and understand or gain greater clarity on some of the emotions with the coping strategies that help you walk through a process. I think just identifying and parsing out your feelings." (P15, user)
- "Favorite feature, definitely the meditation and breathing exercises. I've been really stressed recently, and I've just been working on relaxation, meditation, so I really like those two of the exercises. They've been helping me." (P19, user)
- "I wasn't really that focused on looking at the coping things, because I had a friend who had cancer when she was younger, so she was a big part of asking for coping advice." (P2, nonuser)
- "I just didn't feel the need to [use coping exercises] because I have people to talk to and I feel like I've naturally been able to cope with it independently and with some help. I didn't think I was going to get anything from it, to be honest." (P8, nonuser)
- "I liked the coping strategies section too. I didn't use it very much, but it seemed very straightforward and helpful for people who did need to use it....I'm kind of familiar with some of the coping skills. I see a therapist and also, I'm a psych student, and so I don't think I needed them." (P10, nonuser)

Low User Engagement

- "I think one of the tough parts is that, compared to other apps, there isn't a lot of people....It's tough to go frequently on the app if there's not as many people." (P3)
- "I forget to use it most of the time....This community now is pretty small so it would be great if the community is a little larger." (P4)
- "I think it would be better if there were more people on it at a time. I don't think there was many, I know you guys are just trying this out so it's not like there's as many people on it as I think would eventually be on the app." (P17)
- "I kind of like went in and out. There would be sometimes where I would log on almost daily and then I would go weeks and not look at it. I just thought it was kind of sporadic, but I will say that that is the way that I am with most social media." (P20)
- "I have been meeting with multiple therapists during this whole time and so I was having my needs met by my therapist and the app became a side thing." (P22)

Theme 1: Social Support and Feelings of Isolation

The majority of patients indicated the community feed as their favorite feature, whereas only a minority named the coping exercises. Specifically, patients appreciated the opportunity to connect with peers, receive support, and read about their experiences. Patients felt less isolated knowing and seeing other people of their age dealing with cancer and reporting similar experiences. For example, one patient explained:

> *I thought it was very helpful because I was able to connect and message a few people who went through kind of the exact thing. So, it was a support that I didn't get anywhere else because they had the advice that I needed that nobody else could give me.* [P10]

Theme 2: Use Versus Nonuse of Coping Strategies

Interestingly, several patients noted that despite not using the app intensively, they felt that they still received what they needed from it. For example, one patient said:

> *I liked that it was flexible. That you didn't have to be on it all the time, that you could just check back in on it based on what you wanted to do.* [P17]

Most patients confirmed that they did not use the coping exercises that much, with many of them explaining that they already felt their needs in this area were being met. For those who engaged with this part of the app, the feedback was positive. For example:

> *A lot of the exercises I felt calmed me down, so that worked for me. They were pretty well done.* [P9]

Theme 3: Low User Engagement

Importantly, several participants noted that there were not many users on the app and not many people actively posted content, making the app less engaging and not offering as much opportunity for social connections. Ironically, many of these patients also confessed that they themselves were “more of a viewer than someone to add to things” (P24) on social media and that they were also using the *iaya* app that way. Patients reported varying preferences about the timing of the app in relation to their disease trajectory, with some stating that they found it most useful during active treatment, whereas others said they felt that time was too overwhelming and they preferred exploring the app further after completion of cancer treatment. Several patients suggested that a larger userbase and more turnover of information on the app would facilitate better engagement. However, a few patients also stated that they were hesitant to share content on the app because they did not want to burden other patients with their struggles. Finally, some participants expressed that they would appreciate connecting with patients with a similar diagnosis, because of the variation in treatment for different cancer types, or in a similar life stage, as there are still significant differences between young adults who are college students and those who are young professionals and/or parents.

When specifically asked about the potential negative emotional impact of using the app, only one patient said they felt some of the content shared by other participants triggered certain negative emotions. In addition, a few patients noted that it was sometimes difficult for them to read some of the content on the community feed, particularly when other patients were not doing as well as them. For example, one patient said:

> *Cancer isn’t always that nice to everybody, so then you go in the app and people have had really bad setbacks and so when you’re not having a great emotional day with your own that can be hard to read.* [P20]

Patient-Reported Outcomes

Table 2 shows the baseline and 12-week scores for the patient-reported outcomes. The average CBI-B score among our sample was 84.88 (SD 14.66) at baseline and decreased to 75.74 (SD 14.43) at follow-up, which indicates relatively high self-efficacy for coping with cancer. For the PROMIS self-efficacy for managing emotions, the average T-score was 49.09 (SD 8.21) at baseline and 48.64 (SD 7.80) at follow-up, indicating that our sample had similar self-efficacy for managing emotions compared with the general chronic condition population. For the PROMIS emotional support, the average T-score was 55.24 (SD 7.11) at baseline and 55.37 (SD 8.22) at follow-up, indicating that participants scored half an SD above the mean of the US norm population (T-score: mean 50, SD 10). The mean FACT-G quality of life scores among participants was 75.48 (SD 13.89) at baseline and slightly reduced to 69.91 (SD 15.15) at follow-up. Although these scores are slightly lower than the mean among the norm population, both are within 1 SD of the mean.

Table 2. Mean scores for patient-reported outcomes at baseline and 12 weeks.

Outcome	Baseline (N=25), mean (SD)	12 weeks (n=23), mean (SD)
Self-efficacy for coping with cancer (CBI-B[a])	84.44 (14.66)	75.74 (14.43)
Quality of life (FACT-G[b])	75.48 (13.89)	69.91 (15.15)
Self-efficacy for managing emotions (PROMIS-SE[c]), T-score[d]	49.09 (8.21)	48.64 (7.80)
Perceived emotional support (PROMIS-ES[e]), T-score	55.24 (7.11)	55.37 (8.22)

[a] CBI-B: Cancer Behavior Inventory-Brief Version.

[b] FACT-G: Functional Assessment of Cancer Treatment-General.

[c] PROMIS-SE: Patient-Reported Outcomes Measurement Information System Self-Efficacy for Managing Emotions-Short Form.

[d] For T-scores: in the US norm population, the mean T-score is 50 (SD 10). Higher T-scores represent more of the concept being measured.

[e] PROMIS-ES: Patient-Reported Outcomes Measurement Information System Emotional Support-Short Form.

Discussion

Principal Findings

We developed and pilot-tested an intervention to facilitate an engaged community and learn coping skills among young adults with cancer. In this study, though the intervention was deemed acceptable and overall rated positively by participants, we failed to demonstrate feasibility of the *iaya* app intervention in its present form based on the number of log-ins and coping exercises used. Importantly, our feasibility criteria did not include engagement with the community feed, which was cited by participants as the most interesting and used feature of the app. Furthermore, although self-efficacy for coping with cancer and overall quality of life slightly decreased over time, scores for self-efficacy with regard to emotions and perceived emotional support remained largely stable. Overall, scores were relatively high for all patient-reported outcomes, reflecting the nonclinical nature of our sample.

The finding that user engagement was low, with only 9% (2/23) of participants meeting our combined feasibility criteria of log-ins and coping exercises used, is surprising in the context of ubiquitous smartphone use among the overall young adult population [12,13]. However, selective use and poor retention have been documented for mental health apps, with only a small portion of users using the apps for a long period [35,36]. Alternatively, our a priori feasibility criteria may not have been sufficiently comprehensive, as we did not include all features of app engagement in our definition. At this early stage, we were mostly interested in studying the overall feasibility of implementing a newly developed smartphone app intervention among a novel population. Defining meaningful engagement, or even feasibility, remains a challenge for mental health apps, which has also been described by other investigators, and there is a need to better define and measure engagement in these apps [37].

At least two characteristics of our study may explain why we found low use of our specific app intervention for log-ins and coping exercises used. First, we did not screen for unmet needs in the areas of coping with cancer or managing emotions, and our baseline data indicate that our sample scored relatively well on these domains, which may explain the lack of engagement with coping exercises in our nonclinical sample. Indeed, in the qualitative exit interviews, several patients stated that they did not use the coping exercises section because they felt they either did not need it or had other resources to help them cope with their cancer. In future studies, researchers should consider using the CBI-B as a screener to identify patients who need psychosocial services [28]. Second, our intervention was a stand-alone app with no external factors to motivate young adults with cancer to use it, except for phone reminders early in the first 3 weeks of enrollment and a midstudy call at 6 weeks. Although patients had the option to enable or disable app notifications, in exit interviews, the majority indicated turning off notifications per their personal preference. Although the app was not designed to be integrated within their larger clinical care, some patients indicated that they may have used it more frequently if a clinician encouraged them to use different features of the app or recommended specific elements. In line with the findings from a recent meta-analysis, it seems that despite the potential of apps, using smartphone apps as stand-alone psychological interventions cannot be recommended yet based on the current level of evidence [38]. Therefore, future studies should consider giving participants more directions on how to use the coping exercises and/or integrate the app within their larger clinical care.

Notwithstanding the low use, patients liked the app, deeming it acceptable and easy to use, and they only observed a few technical issues. The main suggestion for further improvement of the app was to increase the number of active users to facilitate greater community engagement and to update content on the app more frequently. We tried to mitigate low user engagement by soliciting *super users* to frequently post content and respond to other users' contributions. In addition, we tried to build an active young adult community by recruiting patients to the app as part of clinical care 6 months before opening our research study for accrual. Between April and November 2019, we enrolled 104 patients outside of the research study. During the 5-month study enrollment period, another 107 patients downloaded the app outside of the research study, resulting in a total of 236 users. However, the number of contributing users overall was still relatively low for a microsocial network app, and many of the research participants who commented on the low user engagement on the app also admitted to adopting more of an observer role on the app, which may also generalize to the larger user base.

At the same time, making the app available before starting our research study presented us with a challenge, as our institutional review board required that we explicitly inform potential study participants that the app was freely available to them outside of the research study, disincentivizing study participation. This also contributed to our low enrollment rate, as 18% (12/65) of approached patients who declined study participation said they would rather download the app separate from the study. One potential solution would be to open up the app to a larger population, including other cancer centers in the Northeast region of the United States, or even nationwide. This would also increase the opportunities for patients to connect with a peer who has been diagnosed with a similar diagnosis, a desire that came up several times during the exit interviews and has been reported previously [16]. This need is understandable and also challenging given the rarity of many cancer diagnoses among young adults.

Challenges

We encountered several challenges in this pilot feasibility study. First, one of our participants dropped out in the first 3 weeks of the study and did not engage in the study procedures. Second, one participant had an older version of the smartphone app installed, which was incompatible with Mixpanel. In addition, if study participants got a new phone during the study period, their earlier use data had to be merged with a new Mixpanel user ID; thus, we had to actively inquire whether participants got a new phone. Third, because of Health Insurance Portability and Accountability Act regulations, the app required participants to log in with their username and password each time they opened the app, even if they briefly switched to another app or function on their smartphone. These clustered log-ins were counted as 1 individual app session. Fourth, although we planned to conduct reminder phone calls to those patients who had not used the app for 14 consecutive days, we were unable to analyze and act upon these data in real time. Therefore, we changed this procedure to a midstudy reminder phone call at 6 weeks. Finally, 5 participants completed the study surveys more than 14 days after completion of their 12-week study period (ie, between 17 and 24 days).

Limitations

Our study had some limitations. This was a pilot study conducted at a single site. Recruitment was suspended 2 months earlier because of the COVID-19 pandemic, and we had a slightly smaller sample size than anticipated. Although we enrolled a clinically diverse sample of 25 patients with different diagnoses, most were well educated, and our sample lacked racial and ethnic diversity. Importantly, as a feasibility pilot, the study did not have a control arm, and thus, we were unable to test the efficacy of the intervention.

Conclusions

In conclusion, although considered acceptable, the *iaya* app did not meet the feasibility criteria we had originally posited, and user engagement was generally low. We gained valuable insights from our qualitative interviews and learned that participants highly valued the community feed aspect, which was not part of our a priori feasibility criteria. Future studies should consider targeting a clinical population with unmet needs, who may benefit more from the coping skills feature. The smartphone app intervention in its present form requires further adaption and refinement before conducting a larger, multisite, randomized clinical trial to assess the efficacy of the *iaya* app intervention on young adults' self-efficacy for coping with cancer and managing emotions, perceived emotional support, and quality of life.

Acknowledgments

This work was supported by a grant from the Oak Foundation and the Elizabeth Stone Endowment Fund at the Dana-Farber Cancer Institute.

Conflicts of Interest

None declared.

References

1. Zebrack BJ. Psychological, social, and behavioral issues for young adults with cancer. Cancer 2011 May 15;117(10 Suppl):2289-2294 [FREE Full text] [doi: 10.1002/cncr.26056] [Medline: 21523748]
2. Zebrack BJ, Corbett V, Embry L, Aguilar C, Meeske KA, Hayes-Lattin B, et al. Psychological distress and unsatisfied need for psychosocial support in adolescent and young adult cancer patients during the first year following diagnosis. Psychooncology 2014 Nov;23(11):1267-1275. [doi: 10.1002/pon.3533] [Medline: 24664958]
3. Roberts CS, Turney ME, Knowles AM. Psychosocial issues of adolescents with cancer. Soc Work Health Care 1998 Jun 29;27(4):3-18. [doi: 10.1300/j010v27n04_02]
4. Burgoyne MJ, Bingen K, Leuck J, Dasgupta M, Ryan P, Hoffmann RG. Cancer-related distress in young adults compared to middle-aged and senior adults. J Adolesc Young Adult Oncol 2015 Jun;4(2):56-63. [doi: 10.1089/jayao.2014.0005] [Medline: 26812552]
5. D'Agostino NM, Penney A, Zebrack B. Providing developmentally appropriate psychosocial care to adolescent and young adult cancer survivors. Cancer 2011 May 15;117(10 Suppl):2329-2334 [FREE Full text] [doi: 10.1002/cncr.26043] [Medline: 21523754]
6. Dyson GJ, Thompson K, Palmer S, Thomas DM, Schofield P. The relationship between unmet needs and distress amongst young people with cancer. Support Care Cancer 2012 Jan 11;20(1):75-85 [FREE Full text] [doi: 10.1007/s00520-010-1059-7] [Medline: 21311915]
7. Forsythe LP, Kent EE, Weaver KE, Buchanan N, Hawkins NA, Rodriguez JL, et al. Receipt of psychosocial care among cancer survivors in the United States. J Clin Oncol 2013 Jun 01;31(16):1961-1969. [doi: 10.1200/jco.2012.46.2101]
8. Olsson CA, Boyce MF, Toumbourou JW, Sawyer SM. The role of peer support in facilitating psychosocial adjustment to chronic illness in adolescence. Clin Child Psychol Psychiatry 2016 Jul 26;10(1):78-87. [doi: 10.1177/1359104505048793]
9. Rabin C, Simpson N, Morrow K, Pinto B. Intervention format and delivery preferences among young adult cancer survivors. Int J Behav Med 2013 Jun;20(2):304-310. [doi: 10.1007/s12529-012-9227-4] [Medline: 22328444]
10. Trevino KM, Fasciano K, Block S, Prigerson HG. Correlates of social support in young adults with advanced cancer. Support Care Cancer 2013 Feb 13;21(2):421-429 [FREE Full text] [doi: 10.1007/s00520-012-1536-2] [Medline: 22790223]
11. Holland LR, Walker R, Henney R, Cashion CE, Bradford NK. Adolescents and young adults with cancer: barriers in access to psychosocial support. J Adolesc Young Adult Oncol 2021 Feb 02;10(1):46-55. [doi: 10.1089/jayao.2020.0027] [Medline: 32486879]
12. Vogels E. Millennials stand out for their technology use, but older generations also embrace digital life. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2019/09/09/us-generations-technology-use/ [accessed 2021-05-13]
13. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-05-13]
14. Lenhart A, Purcell K, Smith A, Zickuhr K. Pew Internet & American Life Project. 2010. URL: https://eric.ed.gov/?id=ED525056 [accessed 2021-05-13]
15. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244 [FREE Full text] [doi: 10.2196/12244] [Medline: 30563811]
16. Lazard AJ, Saffer AJ, Horrell L, Benedict C, Love B. Peer-to-peer connections: perceptions of a social support app designed for young adults with cancer. Psychooncology 2020 Jan;29(1):173-181. [doi: 10.1002/pon.5220] [Medline: 31483913]
17. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013 Dec;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
18. Wesley KM, Fizur PJ. A review of mobile applications to help adolescent and young adult cancer patients. Adolesc Health Med Ther 2015 Aug;6:141-148 [FREE Full text] [doi: 10.2147/AHMT.S69209] [Medline: 26316835]

19. Adam R, McMichael D, Powell D, Murchie P. Publicly available apps for cancer survivors: a scoping review. BMJ Open 2019 Sep 30;9(9):e032510 [FREE Full text] [doi: 10.1136/bmjopen-2019-032510] [Medline: 31575584]
20. Jongerius C, Russo S, Mazzocco K, Pravettoni G. Research-tested mobile apps for breast cancer care: systematic review. JMIR Mhealth Uhealth 2019 Feb 11;7(2):e10930 [FREE Full text] [doi: 10.2196/10930] [Medline: 30741644]
21. Antonovsky A. Health, stress and coping. In: The Jossey-Bass Social and Behavioral Science Series. California, United States: Jossey-Bass Inc; 1979:1-270.
22. Antonovsky A. Unraveling the Mystery of Health: How People Manage Stress and Stay Well. San Francisco: Jossey-Bass; 1987:1-238.
23. Folkman S, Moskowitz JT. Stress, appraisal, and coping. In: Anderson NB, editor. Encyclopedia of Health and Behavior. Thousand Oaks, California, United States: SAGE Publications Inc; 2004:1-1016.
24. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [doi: 10.1037/0033-2909.98.2.310]
25. Cohen S, Syme S. Social Support and Health. Cambridge, Massachusetts, United States: Academic Press; 1985:1-390.
26. Brooke J. SUS - A quick and dirty usability scale. Industrial Usability Evaluation. 1996. URL: https://hell.meiert.org/core/pdf/sus.pdf [accessed 2021-05-13]
27. Merluzzi TV, Philip EJ, Ruhf CA, Liu H, Yang M, Conley CC. Self-efficacy for coping with cancer: revision of the Cancer Behavior Inventory (Version 3.0). Psychol Assess 2018 Dec;30(4):486-499 [FREE Full text] [doi: 10.1037/pas0000483] [Medline: 28504538]
28. Heitzmann CA, Merluzzi TV, Jean-Pierre P, Roscoe JA, Kirsh KL, Passik SD. Assessing self-efficacy for coping with cancer: development and psychometric analysis of the brief version of the Cancer Behavior Inventory (CBI-B). Psychooncology 2011 Mar;20(3):302-312. [doi: 10.1002/pon.1735] [Medline: 20878830]
29. Lee MJ, Romero S, Velozo CA, Gruber-Baldini AL, Shulman LM. Multidimensionality of the PROMIS self-efficacy measure for managing chronic conditions. Qual Life Res 2019 Jun 26;28(6):1595-1603 [FREE Full text] [doi: 10.1007/s11136-019-02116-w] [Medline: 30806873]
30. Gruber-Baldini AL, Velozo C, Romero S, Shulman LM. Validation of the PROMIS measures of self-efficacy for managing chronic conditions. Qual Life Res 2017 Jul 26;26(7):1915-1924 [FREE Full text] [doi: 10.1007/s11136-017-1527-3] [Medline: 28239781]
31. Hahn EA, DeWalt DA, Bode RK, Garcia SF, DeVellis RF, Correia H, PROMIS Cooperative Group. New English and Spanish social health measures will facilitate evaluating health determinants. Health Psychol 2014 May;33(5):490-499 [FREE Full text] [doi: 10.1037/hea0000055] [Medline: 24447188]
32. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. J Am Med Assoc 2013 Feb 27;309(8):814-822. [doi: 10.1001/jama.2013.879] [Medline: 23443445]
33. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy Scale: development and validation of the general measure. J Clin Oncol 1993 Mar;11(3):570-579. [doi: 10.1200/JCO.1993.11.3.570] [Medline: 8445433]
34. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
35. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
36. Bauer M, Glenn T, Geddes J, Gitlin M, Grof P, Kessing LV, et al. Smartphones in mental health: a critical review of background issues, current status and future concerns. Int J Bipolar Disord 2020 Jan 10;8(1):2 [FREE Full text] [doi: 10.1186/s40345-019-0164-x] [Medline: 31919635]
37. Torous J, Michalak EE, O'Brien HL. Digital health and engagement-looking behind the measures and methods. JAMA Netw Open 2020 Jul 01;3(7):e2010918 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]
38. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 2;2(1):118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]

Abbreviations

CBI-B: Cancer Behavior Inventory-Brief Version
FACT-G: Functional Assessment of Cancer Therapy-General
PROMIS: Patient-Reported Outcomes Measurement Information System
REDCap: Research Electronic Data Capture

Edited by R Kukafka; submitted 16.10.20; peer-reviewed by L Horrell, L Jibb; comments to author 13.02.21; revised version received 04.03.21; accepted 25.04.21; published 11.06.21.

Please cite as:
Poort H, Ryan A, MacDougall K, Malinowski P, MacDonald A, Markin Z, Pirl W, Greer J, Fasciano K
Feasibility and Acceptability of a Mobile Phone App Intervention for Coping With Cancer as a Young Adult: Pilot Trial and Thematic Analysis
J Med Internet Res 2021;23(6):e25069
URL: https://www.jmir.org/2021/6/e25069
doi:10.2196/25069
PMID:34114957

Original Paper

Preferences for mHealth Technology and Text Messaging Communication in Patients With Type 2 Diabetes: Qualitative Interview Study

Julie C Lauffenburger[1], PharmD, PhD; Renee A Barlev[1], MD, MPH; Ellen S Sears[1], BS; Punam A Keller[2], PhD; Marie E McDonnell[1], MD; Elad Yom-Tov[3], PhD; Constance P Fontanet[1], MPH; Kaitlin Hanken[1], MPH; Nancy Haff[1], MD, MPH; Niteesh K Choudhry[1], MD, PhD

[1]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States
[2]Dartmouth College, Dartmouth, NH, United States
[3]Microsoft Research, Herzliya, Israel

Corresponding Author:
Julie C Lauffenburger, PharmD, PhD
Brigham and Women's Hospital, Harvard Medical School
1620 Tremont Street
Suite 3030
Boston, MA, 02120
United States
Phone: 1 6175258865
Email: jlauffenburger@bwh.harvard.edu

Abstract

Background: Individuals with diabetes need regular support to help them manage their diabetes on their own, ideally delivered via mechanisms that they already use, such as their mobile phones. One reason for the modest effectiveness of prior technology-based interventions may be that the patient perspective has been insufficiently incorporated.

Objective: This study aims to understand patients' preferences for mobile health (mHealth) technology and how that technology can be integrated into patients' routines, especially with regard to medication use.

Methods: We conducted semistructured qualitative individual interviews with patients with type 2 diabetes from an urban health care system to elicit and explore their perspectives on diabetes medication–taking behaviors, daily patterns of using mobile technology, use of mHealth technology for diabetes care, acceptability of text messages to support medication adherence, and preferred framing of information within text messages to support diabetes care. The interviews were digitally recorded and transcribed. The data were analyzed using codes developed by the study team to generate themes, with representative quotations selected as illustrations.

Results: We conducted interviews with 20 participants, of whom 12 (60%) were female and 9 (45%) were White; in addition, the participants' mean glycated hemoglobin A_{1c} control was 7.8 (SD 1.1). Overall, 5 key themes were identified: patients try to incorporate *cues* into their routines to help them with consistent medication taking; many patients leverage some form of technology as a cue to support adherence to medication taking and diabetes self-management behaviors; patients value simplicity and integration of technology solutions used for diabetes care, managing medications, and communicating with health care providers; some patients express reluctance to rely on mobile technology for these diabetes care behaviors; and patients believe they prefer positively framed communication, but communication preferences are highly individualized.

Conclusions: The participants expressed some hesitation about using mobile technology in supporting diabetes self-management but have largely incorporated it or are open to incorporating it as a cue to make medication taking more automatic and less burdensome. When using technology to support diabetes self-management, participants exhibited individualized preferences, but overall, they preferred simple and positively framed communication. mHealth interventions may be improved by focusing on integrating them easily into daily routines and increasing the customization of content.

(J Med Internet Res 2021;23(6):e25958) doi:10.2196/25958

KEYWORDS

diabetes; technology; mobile health; medication adherence; mobile phone

Introduction

Background

Achieving optimal glycemic control substantially decreases the risk of long-term complications in individuals with type 2 diabetes [1-3]. Yet, more than 40% of patients with diabetes do not achieve their glycemic targets [4]. Medication adherence is an important factor in achieving diabetes control and improving clinical outcomes [5-8]. On average, less than half of patients take their medications as prescribed [7,8]. Managing diabetes also requires simultaneous adherence to diet and physical activity goals and achieving weight loss [5,9-12]. Although health care professionals can recommend treatments and promote healthy behaviors, activities like medication adherence, weight monitoring, dietary choices, and exercise must be self-managed by the patients [13,14]. Accordingly, to meaningfully improve outcomes, individuals with diabetes need regular support and feedback to help them manage these behaviors on their own, ideally delivered via mechanisms that they already use.

As of 2019, more than 80% of the US population owns smartphones, with relatively consistent uptake across most sociodemographic groups. Among the many functionalities of smartphones, text messaging is especially inexpensive and can provide reminders, education, and motivational support on an ongoing basis [15]. Growing evidence supports the effectiveness of text messaging–based approaches for diabetes self-management [15,16], although the magnitude of benefit that most patients receive is relatively modest [16-18].

One reason for this limited effectiveness of text messaging–based interventions in patients with diabetes may be that patients' perspectives have been insufficiently incorporated into the message content and delivery. Little is known about specific preferences for mobile health (mHealth) technology for diabetes and how the technology could integrate into patients' routines, especially with regard to medication use [19-21]. Patients may also exhibit specific preferences for how they receive text messaging–based communication from health care professionals, which may affect patients' diabetes self-management activities.

Objective

Therefore, we conducted in-depth, semistructured qualitative interviews with individuals with type 2 diabetes to elicit and explore their perspectives on (1) the medication-taking behaviors and challenges in diabetes; (2) their daily patterns of using mobile technology; (3) their use of mHealth technology for diabetes care, medication management, and provider communication; (4) the acceptability of a text messaging–based system for supporting medication adherence; and (5) their preferred framing of information within text messages to support diabetes care.

Methods

Study Participants and Recruitment

Participants were eligible if they were 18 years or older, diagnosed with type 2 diabetes, and were taking 1 or more prescribed oral medication for that condition. It should be noted that participants who were also using insulin were not explicitly excluded to enhance generalizability. We recruited a purposive sample of participants for qualitative interviews through direct clinician referral from Mass General Brigham health centers (formerly Partners Health care) as well as posting in and recruiting from a web-based database of participants who expressed an interest in participating in research studies in and around Boston, Massachusetts [22]. The recruitment and interviews occurred between January and March 2020. Informed consent was obtained from all interview participants, and they received a US $30 honorarium on the completion of the interview.

In this study, we followed the established standards for reporting qualitative research (ie, the COREQ [Consolidated Criteria for Reporting Qualitative Research] checklist) to ensure conformity with qualitative research standards, as recommended by the EQUATOR (Enhancing the Quality and Transparency of Health Research) Network [23,24]. Throughout the paper, we have referred to participants by their study ID to preserve anonymity. The Mass General Brigham Review Board approved this study.

Qualitative Interviews

To elicit deeply personal accounts and develop a greater understanding of participants' perceptions of diabetes management and the role of technology and text messaging–based solutions in their care, we chose to use semistructured, one-on-one qualitative interviews. To do so, the lead author, who had experience in qualitative methods, medication adherence, and clinical pharmacy, first drafted a comprehensive semistructured interview guide and circulated it for feedback and iterative refinement from other coinvestigators with expertise in qualitative methods, clinical medicine, mobile phone technology, diabetes, and medication adherence. Nonparticipant volunteers also pilot tested the guide, which led to no substantial changes. The guide covers separate but overlapping topics: (1) medication-taking behaviors and challenges in diabetes; (2) their use of mobile technology in their daily lives; (3) their use of technology for diabetes care, medication management, and provider communication; (4) general acceptability of a nonspecific text messaging–based system for medication management; and (5) preferred framing of information within text messages to support diabetes care (Multimedia Appendix 1, Table S1). The guide was used flexibly during the interviews to follow the natural flow of conversation and allow for free-flowing discussion by participants.

The participants were also shown 15 example text messages to respond to toward the end of the interview to further elicit perspectives on the framing of information. These text messages have been used in prior text messaging–based interventions for diabetes (Multimedia Appendix 1, Table S2) [25,26].

All interviews were conducted in English by a trained moderator who is also a practicing pharmacist (JCL). The interviews were conducted in person at secure locations around the Brigham and Women's Hospital. At the beginning of the interview, the participants completed a baseline questionnaire that included questions about self-reported adherence to medication [27,28] and patient activation [29,30] (Multimedia Appendix 1, Figure S1). Each interview lasted between 30 and 75 minutes (mean 38.5, SD 10.3), and a sequential identification number was assigned to each interview. We conducted interviews until saturation was reached; saturation of the data was reached by patient 19 and reaffirmed with patient 20.

Data Analyses

The audio-recorded interviews were transcribed verbatim. Transcripts were checked for accuracy against the recordings. To conduct the data analysis, we used immersion-crystallization methods [31,32]. In brief, this approach involves immersing in the data and then crystalizing salient themes during this process. Three investigators independently annotated a selection of transcripts and devised preliminary topic codes, with major categories consisting of technology, health, medication, routine, and texting. Each transcript was analyzed by at least two reviewers. Preliminary coding revealed themes around (1) having cues in medication regimen, (2) preferences for simplicity in communication, (3) concern about the applicability of communication, (4) reluctance to rely on technology, and (5) preferences for positively framed texts. After discussion and review with the coinvestigators, the preliminary codes were revised and agreed upon. We also identified broad themes at this stage.

Dedoose software version 8.3.10 (SocioCultural Research Consultants, LLC) was used for storage, handling, and analysis of the data set. We continued the immersion-crystallization approach until all the data were examined. Several early themes were clear: the role of habits in medication management, tendencies to use technology for managing diabetes but also being reluctant to rely on it, the perceived need for technology to seamlessly integrate into routines to be useful, and preferences for positively framed health communication. After coding, all the transcripts were then reread to identify any additional themes.

Results

Overview

We conducted interviews with 20 participants, and their key baseline sociodemographic and clinical characteristics are shown in Table 1. Among the participants, 60% (12/20) were female and 45% (9/20) were White; in addition, they had a mean glycated hemoglobin A_{1c} of 7.8 (SD 1.1). Although 65% (13/20) of the participants reported missing at least one day of medication in the last 30 days, the overall rates of adherence were modestly high (ie, a mean of 1.15 days missed, SD 1.1) [27,28].

In the interviews, participants reflected in detail about their medication-taking behaviors and routines; how they integrate mobile technology in their daily lives; their experiences with using mHealth technology to manage their diabetes; and their preferred method of receipt and framing of health-related communication, especially with regard to their medications.

On the basis of these interviews, we identified 5 key themes related to medication taking, technology, and preferences, which are summarized in Textbox 1. We present each of these themes in more detail in the following sections, with representative quotations by participants from the transcripts shown in italics. We also share the specific reactions to potential text messages shown to the participants.

Table 1. Participant characteristics.

Participant ID	Age (years)	Gender	Race and ethnicity	Education level	Medications in regimen, n	Latest glycated hemoglobin A_{1c} value	Days that medications were missed in the last 30 days, n
P1	57	Male	White	Some college	15	7.2	1
P2	36	Female	Black, Latino	Some college	3	9.2	2
P3	42	Female	White	Some college	3	6.4	0
P4	68	Male	White	College graduate	6	—[a]	1
P5	65	Male	Black	Some college	7	—[a]	2
P6	54	Female	Black	College graduate	4	9.1	0
P7	56	Male	White	Some college	10	9.7	2
P8	56	Female	Asian	Postgraduate	14	6.4	2
P9	55	Female	Black	College graduate	2	7.3	0
P10	48	Female	Black	Some college	14	6.9	3
P11	45	Female	White	College graduate	4	6.8	0
P12	54	Male	White	Postgraduate	9	7.1	1
P13	78	Male	White	College graduate	8	9.3	1
P14	63	Male	Black, native American	Postgraduate	7	7.3	1
P15	71	Male	White	Postgraduate	8	8.6	1
P16	56	Female	Black	Some college	5	8.0	0
P17	59	Female	Latino	Some college	3	7.1	0
P18	21	Female	Latino	Some college	1	7.0	3
P19	58	Female	Black	Some college	14	6.9	0
P20	47	Female	White	Postgraduate	3	8.0	3

[a] Data not reported.

Textbox 1. Summary of key themes.

Themes and key takeaways

1. Patients try to incorporate cues into their routines to help them with consistent medication taking:
    - Value of establishing a daily routine in supporting adherence
    - Primarily use sight cues or habit cues
2. Many patients leverage some form of technology as a cue to support adherence to medication taking and diabetes self-management behaviors:
    - Mobile phones serve as sight or habit cue for medication taking, either deliberately or accidentally
3. Value of simplicity and integration in technology solutions used for diabetes care, managing medications, and communicating with health care providers:
    - Simplicity in communication and integration into routine can prevent exhaustion from being connected
    - Straightforward, direct communication is preferred
4. Some patients express a reluctance to rely on mobile technology for these diabetes care behaviors:
    - Concern about integration of mobile phones into daily lives for diabetes because of obsession with control values or concerns about screen time
5. Patients believe they prefer positively framed communication, but communication preferences are highly individualized:
    - General preference for positively framed information
    - Concerns about relevance of information to themselves or others, including advice and describing social support

Patients Try to Incorporate Cues Into Their Routines to Help Them With Consistent Medication Taking

Participants described in great detail the value of a routine in supporting daily adherence to their diabetes medications, such as maintaining specific wake-up times, integrating medication taking into their bathroom use, or eating breakfast at the same time as their medications. Their medication-taking routines are most commonly supported by event-based cues centered on activities that are part of their normal routine, such as drinking coffee or using the bathroom or their mobile phone. These were sometimes supported by visual cues, such as sticky notes or a colorful pillbox in an obvious location. Others describe the act of glucose testing itself as supporting their medication taking and, therefore, the coupling of several activities that contribute to better blood sugar control. Conversely, for medications intended to be taken once a week, participants commonly described temporal cues, such as the day of the week:

> *If I remember church, I can remember to take the medication on Sundays.* [P16]
>
> *I like little hacks; trying to remember little hacks like by the coffee pot or maybe in the restroom in the morning. Where you're brushing your teeth and you see your package of pills.* [P4]
>
> *Even if my day is different—like I'm out all day today, so I'll probably have dinner before I go home, but as soon as I go home my pills are on my computer, which is one of the first things I do, so I take 'em right away.* [P1]
>
> *My container is sitting right there [by the keys on the kitchen counter], so there's a good chance I'm gonna look at it before I walk out the door. That's the strategy.* [P12]
>
> *It's automatic, almost. It's almost part of getting my first cup of coffee. I put my machine in, press it to get the coffee, and then while it's coming, I'll go over and test.* [P15]
>
> *I'm programmed. I wake up. I make sure my coffee is brewing and then the pill before the coffee. That's like a religion for me.* [P17]

Participants who largely reported a less *automatic* routine struggled with remembering to take their medications, filling their pillboxes, or recalling later whether they actually took their medications or not. Weekends appeared particularly difficult for participants because of the break in routine. Relatively few participants described behaviors that were fully automatic:

> *It's not that I don't remember the box. It's that I don't remember to fill certain medications in the box.* [P8]
>
> *The weekend is the hardest because in the weekend, routines change a little. I'm not rushing.* [P17]
>
> *Sometimes I forget, I get confused, and I don't remember that I have already taken it or not. If I take another one, then I feel worse because my sugar levels are really low. I think that's the hard part—to remember if I had it or not.* [P18]
>
> *I'm usually doing other things while I'm trying to remember the medication and get the kids ready. "Did I take the medication? I don't remember." Sometimes I didn't even realize I missed one.* [P20]

Many Patients Leverage Some Form of Technology as a Cue to Support Adherence to Medication Taking and Diabetes Self-management Behaviors

For about half of the participants, their mobile phones served as cues for medication taking. In some cases, participants generated specific cues using their phones as reminders, such as alarms or calendar reminders. In other cases, participants described physically storing their medications near their phone during the day because they looked at their phones frequently and therefore used them as direct cues. Others recognized that using their phones or other technology would be difficult to support medication taking if they were not seamlessly integrated into their daily lives. Whether participants deliberately used their phones for medication taking appeared to depend in part on how strongly they felt they needed that as a cue or how strongly they felt most generally about their need for cues for medication taking:

> *The eight AM one I have missed. I usually have it in my pocket for the most part. The 12:00 one I don't miss 'cause it's usually in this pocket with the phone, and this phone comes in and out of my pocket all day long.* [P1]
>
> *I always set two [alarms] to take medication; usually if I took it, I uncheck the other one so the alarm doesn't go off. If it does go off, that means I didn't take it. I have to do it, otherwise there's no way I'm gonna remember.* [P2]

P8 explained why a phone may not always work by stating the following:

> *It's easier for men, because they have more pockets. I feel like I don't [notice it]; I have this cross-body thing [bag] that can hold my phone.*

P12 explained why he uses his phone's calendar feature:

> *Because they remind me, and it's something I don't have to store in my head. It's just that some things are better left to a machine and automation.*

P14 stated the following:

> *The cell phone could be a screwdriver, or a hammer. It's a tool. I don't have an emotional attachment to it. Now I'm gonna sound like a boomer. I got along without 'em. To me, it's just a tool, not the center of my life.*

Patients Value Simplicity and Integration for Technology Solutions Used for Diabetes Care, Managing Medications, and Communicating With Health Care Providers

During the interview, participants reiterated that they disliked complicated technology or repetitive communication and offered several personal examples about likes and dislikes of communication that they currently receive. Simplicity was also valued because of its benefits in preventing exhaustion that comes from being constantly connected to technology and using technology for diabetes care. About one-third of the participants stated that they preferred written mobile communication rather than telephone calls, in part because of this desire for straightforwardness:

> *I just want you to get to the point and move on. Especially if you're gonna remind me to take my pills. I mean it's like did you take your pills today? You can either sit there and wait for a yes or a no. it's very simple. Simplicity; I like simplicity.* [P1]
>
> *I love the little idea of—I love how straightforward the Brigham one [text] is. I love that. "Hey, did you take your meds? Respond back yes or no." I think that's great. Just straightforward.* [P3]
>
> *I hate talking on the phone, so if I could have done that on my phone, like in an app or texting, would've been 1000 times—'cause I kept putting it off because I hate talking on the phone. Anything I don't have to do face to face and can actually text or something, way better.* [P11]
>
> *Personally, I feel that watching out for what you have to eat and taking the pills and all this is frustrating in life. The longer the text message is the more like we're like, "Blah." We don't pay attention to it 'cause we don't want more stress. We don't want somebody telling, giving you all this lecture.* [P17]

In another context, when specifically asked about whether they would use text messages for diabetes, the concept was largely received positively, as long as the information delivered was simple and considered actionable. Participants largely discussed managing their diabetes as something within their locus of control, or their own responsibility, and noted that text messages would have to adapt to that. Specific feedback on each example text message shown to the participants is displayed in Multimedia Appendix 1, Table S2:

> *I think that would be short, to the point at the time when you should be takin' it because texts, I think, are close to instantaneous to people. Most people have their phone in their pocket, especially the younger crowd. Older people, maybe not so much, but I think, even my generation is really into the phone, too. We mimic the younger kids because they do all the innovation and then show us how to do it.* [P12]
>
> *The daily stuff, I think I have it pretty much under control. I don't think I need constant reminders, but I can see at some point it probably wouldn't hurt.* [P7]
>
> *I like to text. I like to communicate. Communication is vital in my life and also with my own PCP [primary care physician], my own doctor, and all in your family and all. Yeah, so I like texts.* [P17]

Some Patients Express Reluctance to Rely on Mobile Technology for These Diabetes Care Behaviors

Although participants largely reported high integration of mobile phones into their daily lives, they also expressed concern about this practice. In particular, they felt that technology could lead to an obsession with improving their diabetes numbers because of the rapid acquisition of those data. They described the physical exhaustion that comes from constantly concentrating on disease management, which may be amplified by technology use. There were also some stated concerns about overall screen time that led them to actively ignore or turn off some features that may be necessary to depend on mobile technology for diabetes care. Reluctance to rely on mobile technology may be associated with beliefs that technology is intrusive or patients are relinquishing control of their diabetes self-management:

> *You have to be your own advocate, but it's a full-time job for me to manage my healthcare around my doctors and my prescriptions and, you know, everything. It's just there are days where I just shut it off [the phone] and say, I can't today. I try to limit screen time anyways; my phone will be at the bottom of my bag, and it might go off, and I just won't hear it.* [P8]
>
> *It's almost that I have a love hate relationship with the thing. I love it when I'm low. I hate it when I'm high. I tend to rely on my A_{1C} as my overall measure of how well I'm doing with my diabetes.* [P12]
>
> *My phone is always on silent. I can't stand the noise.* [P11]

Patients Believe They Prefer Positively Framed Communication, but Communication Preferences Are Highly Individualized

Participants expressed strong preferences for positively framed information within the communication they receive from health care providers. In their mind, positively framed communication involved motivational statements rather than penalizing statements or statements about the physical consequences of not taking medications. They also articulated similar desires for text messages to support medication use:

> *Because I'm into positive thinking in the way of trying to stay positive, trying to say things positive, so I like it when people send you those messages.* [P5]
>
> *More of encouragement than, "You missed this. You didn't do this," more of the whole, "Let's get this done. Did you do"—more encouragement than penalized, feeling penalized.* [P7]
>
> *Trying to find something positive, either a way to do something positive or a positive outcome of, "If you do this—"* [P11]

Overall, participants wanted customized communication, and unless the information was applicable to them (eg, if they did not have family members involved in their care or social support), they surmised that they would likely ignore the text. Participants did not like texts that they deemed unrelated to their personal routine, such as texts with tips or advice. When asked about other types of communication preferences, such as whether to provide additional information beyond reminders, such as specific links to more information about diabetes self-care, or mentioning friends or family as sources of social support, participants had varied opinions about whether that would help them support medication taking:

> *People do things different ways. I was getting a text from somebody, I would wanna make it feel personal because then you tend not to ignore it than if it's a generic.* [P9]
>
> *Friends and family's good, but friends and family a lotta times tend to be on the negative side of things and not as encouraging as you'd like, more, "You didn't do it, so you're bad."* [P7]
>
> *To point out that friends or family could help me—but it's, like, "Well, they're not around."* [P3]
>
> *I can't tell my family [about my diabetes] 'cause they're very judge-y. I can't stand when people micro-watch you. I don't feel comfortable telling by family because they'll be, like, "Well, can you eat that? Can you do that?"* [P11]
>
> *It's pretty much known that if you eat well and exercise 30 minutes a day. I don't need to be reminded of that.* [P6]

Discussion

Principal Findings

There is an urgent need to identify effective and sustainable strategies to engage individuals with type 2 diabetes on an ongoing basis and in their usual environments. mHealth technology has the potential to support self-care activities, only if its delivery of communication could be optimized [33]. In this qualitative interview study, participants expressed some hesitation about the role of mobile technology in supporting diabetes self-management but have largely incorporated or are open to incorporating it as a cue in the pursuit of making medication taking more automatic and less burdensome. When using technology to support diabetes self-management, participants generally preferred simple and positively framed communication.

Comparison With Prior Work

Of the few prior studies evaluating preferences for mHealth technology in diabetes, one conducted among 15 participants in the United Kingdom found a similar *love-hate* relationship with the increased awareness of glucose control provided by mHealth [34]. In another study of focus groups with 23 participants in the United Kingdom that concentrated on the acceptability of text messaging, participants relayed similar concerns about the relevance of content and expressed similar reluctance to rely on technology, largely because they felt that they should take personal responsibility for remembering to take medications [20]. Our study builds on this existing research but provides further insight into the role of mHealth in medication taking and preferences regarding message framing. Although preferences for simplicity have been observed in prior studies [35], it is notable that our US participants emphasized its importance in the context of reducing cognitive load and screen time [20,34,36,37].

Our findings also differ somewhat from other prior qualitative studies of text messaging, as our participants placed less emphasis on texts that invoke social concepts. A qualitative interview study in Argentina found that patients appreciated socially framed content in text messages [38], a finding consistent with survey data showing modestly high preferences for texts incorporating social support [21]. These differences could be attributable to culture or preferences for social support across different countries and populations, possibly because culture in the United States is generally more individualistic [39-42].

Related literature beyond text messaging programs about mobile apps reveals findings that are similar to ours. Patients have strong preferences for diabetes apps that are easy and efficient to use [43,44], and strong integration in patients' daily routines has also been noted as desired features of mobile apps [45,46]. Similarly, communication preferences themselves have been thought to be highly individual [47]. However, several differences between mobile apps and text messaging programs should be noted: (1) participants must have a smartphone to use a mobile app, (2) participants must more actively open or engage with apps to derive benefits, and (3) participants express specific financial concerns about apps [44,46]. These considerations may be less relevant to mHealth messaging programs [15,44].

Implications of the Findings

These findings offer several lessons for improving mHealth messaging programs for diabetes. Participants identified several strategies that are consistent with evidence from behavioral science, such as salience and framing of information [48-50]. For example, one key concept from the interviews was the desire for applicability and customization to ensure that the content is relevant and, therefore, salient to patients. Other limited existing evidence also suggests that mobile technology–based interventions may be most effective when information is tailored to the characteristics of individual patients, such as their specific barriers to adherence; of course, not all barriers can be easily addressed in text message solutions [51,52]. Furthermore, participants identified the importance of framing of communication, which is of demonstrated importance in other contexts, such as preventive health screenings and vaccinations [48,49,53]. These interviews suggested that framing, especially positively, may affect influence whether patients respond to text messaging programs in diabetes. Although, on average, patients prefer positive framing, persuasiveness of other framing may depend on patients' specific barriers, and framing in text messaging programs specifically may warrant further empirical exploration.

Other design factors also appear to be important for the success of text messaging–based interventions in diabetes. The simplicity and ease of incorporating a hypothetical text messaging–based system into patients' daily lives appear to be central to engendering the automaticity in medication taking that patients aim for. Mobile phones may already serve as habit-based cues, which could support patients until they no longer need to rely on those cues. However, there may be other ways in which text messaging programs should be better integrated into clinical care, such as more seamless communication with patient portals or electronic health record systems. Overall, technology can overcome many medication adherence barriers, including enhancing planning, being objective or not judgmental, and a ubiquitous presence, but more research is needed on the characteristics of patients who may benefit the most. Similarly, text messages for medication adherence could also be more persuasive than text messages for diet and exercise reminders because adherence behaviors are generally easier to implement; however, any differences are not yet well characterized.

Limitations

This study has several limitations. First, this study was conducted in an urban academic medical center in eastern Massachusetts and recruited directly from clinics or from a web-based database of interested subjects, which could have affected generalizability; however, we enrolled a clinically and demographically diverse sample of participants. Second, the mean age of the study sample was 54 years (SD 12.3), which reflects the age distributions of individuals with type 2 diabetes in the United States; nevertheless, the results could have underrepresented younger viewpoints, who may have differing perspectives on technology use [54]. Third, because we conducted the interviews in person, response bias may have been possible. Finally, although participants using insulin were not excluded, the results may not be generalizable to patients exclusively using injectables to manage their diabetes.

Conclusions

Participants appeared to express some trepidation about the daily role of mobile technology, but they have largely incorporated it or are open to incorporating it as a cue in the pursuit of making medication taking more automatic and less burdensome. mHealth interventions may be improved by focusing on easy integration into daily routines and increasing personalization. Careful, tailored application of behavioral science theories may be especially important in a society that increasingly relies on at-home, virtual care for managing diabetes.

Acknowledgments

This research was supported by a grant from the National Institutes of Health National Institute on Aging to Brigham and Women's Hospital (NKC is Center PI: P30AG064199). JCL was also supported in part by a career development grant (K01HL141538) from the National Institutes of Health.

Conflicts of Interest

RAB is funded by an unrestricted educational grant from Boehringer Ingelheim to the Brigham and Women’s Hospital. NKC has received research grants from Boehringer Ingelheim and Humana payable to Brigham and Women's Hospital and receives consulting fees and holds equity in RxAnte.

Multimedia Appendix 1
Supplementary tables and figure.
[DOCX File , 177 KB - jmir_v23i6e25958_app1.docx]

References

1. Huang ES, Liu JY, Moffet HH, John PM, Karter AJ. Glycemic control, complications, and death in older diabetic patients: the diabetes and aging study. Diabetes Care 2011 Jun 19;34(6):1329-1336 [FREE Full text] [doi: 10.2337/dc10-2377] [Medline: 21505211]
2. Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, et al. Management of hyperglycaemia in type 2 diabetes: a patient-centered approach. Position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia 2012 Jun 20;55(6):1577-1596. [doi: 10.1007/s00125-012-2534-0] [Medline: 22526604]
3. Lau DC, Teoh H. Impact of current and emerging glucose-lowering drugs on body weight in type 2 diabetes. Can J Diabetes 2015 Dec;39 Suppl 5:148-154. [doi: 10.1016/j.jcjd.2015.09.090] [Medline: 26654858]
4. American Diabetes Association. 8. Pharmacologic approaches to glycemic treatment: standards of medical care in diabetes—2018. Diabetes Care 2017 Dec 08;41(Supplement 1):73-85. [doi: 10.2337/dc18-s008]
5. Carls GS, Tuttle E, Tan R, Huynh J, Yee J, Edelman SV, et al. Understanding the gap between efficacy in randomized controlled trials and effectiveness in real-world use of GLP-1 RA and DPP-4 therapies in patients with type 2 diabetes. Diabetes Care 2017 Nov 11;40(11):1469-1478. [doi: 10.2337/dc16-2725] [Medline: 28801475]
6. Fernández A, Quan J, Moffet H, Parker MM, Schillinger D, Karter AJ. Adherence to newly prescribed diabetes medications among insured Latino and White patients with diabetes. JAMA Intern Med 2017 Mar 01;177(3):371-379 [FREE Full text] [doi: 10.1001/jamainternmed.2016.8653] [Medline: 28114642]

7. Lin L, Sun Y, Heng BH, Chew DE, Chong P. Medication adherence and glycemic control among newly diagnosed diabetes patients. BMJ Open Diabetes Res Care 2017 Jul 31;5(1):e000429 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000429] [Medline: 28878942]
8. Capoccia K, Odegard PS, Letassy N. Medication adherence with diabetes medication: a systematic review of the literature. Diabetes Educ 2016 Feb 04;42(1):34-71. [doi: 10.1177/0145721715619038] [Medline: 26637240]
9. Lauffenburger JC, Lewey J, Jan S, Lee J, Ghazinouri R, Choudhry NK. Association of potentially modifiable diabetes care factors with glycemic control in patients with insulin-treated type 2 diabetes. JAMA Netw Open 2020 Jan 03;3(1):e1919645 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.19645] [Medline: 31968115]
10. Lauffenburger JC, Mahesri M, Choudhry NK. Not there yet: using data-driven methods to predict who becomes costly among low-cost patients with type 2 diabetes. BMC Endocr Disord 2020 Aug 17;20(1):125 [FREE Full text] [doi: 10.1186/s12902-020-00609-1] [Medline: 32807156]
11. Gatwood JD, Chisholm-Burns M, Davis R, Thomas F, Potukuchi P, Hung A, et al. Disparities in initial oral antidiabetic medication adherence among veterans with incident diabetes. J Manag Care Spec Pharm 2018 Apr;24(4):379-389. [doi: 10.18553/jmcp.2018.24.4.379] [Medline: 29578849]
12. Mayberry LS, Osborn CY. Family support, medication adherence, and glycemic control among adults with type 2 diabetes. Diabetes Care 2012 Jun 26;35(6):1239-1245 [FREE Full text] [doi: 10.2337/dc11-2103] [Medline: 22538012]
13. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Fischl AH, et al. Diabetes self-management education and support in type 2 diabetes. Diabetes Educ 2017 Feb 24;43(1):40-53. [doi: 10.1177/0145721716689694] [Medline: 28118121]
14. Nadolsky KZ. Re: Consensus statement by the American Association of Clinical Endocrinologists and American College of Endocrinology on the comprehensive type 2 diabetes management algorithm – 2018 executive summary. Endocr Pract 2018 May;24(5):498. [doi: 10.4158/ep-2018-0062]
15. Lauffenburger JC, Choudhry NK. Text messaging and patient engagement in an increasingly mobile world. Circulation 2016 Jan 14;133:555-556. [doi: 10.1161/circulationaha.116.021182]
16. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar 01;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
17. Dobson R, Whittaker R, Dale LP, Maddison R. The effectiveness of text message-based self-management interventions for poorly-controlled diabetes: a systematic review. Digit Health 2017 Nov 09;3:2055207617740315 [FREE Full text] [doi: 10.1177/2055207617740315] [Medline: 29942620]
18. Sahin C, Courtney KL, Naylor P, E Rhodes R. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: a systematic review and a meta-analysis. Digit Health 2019 Apr 22;5:2055207619845279 [FREE Full text] [doi: 10.1177/2055207619845279] [Medline: 31041110]
19. Long H, Bartlett YK, Farmer AJ, French DP. Identifying brief message content for interventions delivered via mobile devices to improve medication adherence in people with type 2 diabetes mellitus: a rapid systematic review. J Med Internet Res 2019 Jan 09;21(1):e10421 [FREE Full text] [doi: 10.2196/10421] [Medline: 30626562]
20. Bartlett YK, Newhouse N, Long HA, Farmer A, French D. What do people with type 2 diabetes want from a brief messaging system to support medication adherence? Patient Prefer Adherence 2019 Sep;Volume 13:1629-1640. [doi: 10.2147/ppa.s217843]
21. Bartlett YK, Farmer A, Rea R, French DP. Use of brief messages based on behavior change techniques to encourage medication adherence in people with type 2 diabetes: developmental studies. J Med Internet Res 2020 May 13;22(5):e15989 [FREE Full text] [doi: 10.2196/15989] [Medline: 32401214]
22. Rally with Mass General Brigham. Mass General Brigham. URL: https://rally.partners.org/ [accessed 2021-01-22]
23. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
24. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research. Acad Med 2014;89(9):1245-1251. [doi: 10.1097/acm.0000000000000388]
25. Lauffenburger JC, Lewey J, Jan S, Makanji S, Ferro CA, Krumme AA, et al. Effectiveness of targeted insulin-adherence interventions for glycemic control using predictive analytics among patients with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2019 Mar 01;2(3):e190657 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0657] [Medline: 30874782]
26. Choudhry NK, Isaac T, Lauffenburger JC, Gopalakrishnan C, Lee M, Vachon A, et al. Effect of a remotely delivered tailored multicomponent approach to enhance medication taking for patients with hyperlipidemia, hypertension, and diabetes: the STIC2IT cluster randomized clinical trial. JAMA Intern Med 2018 Sep 01;178(9):1182-1189 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3189] [Medline: 30083727]
27. Wilson IB, Lee Y, Michaud J, Fowler FJ, Rogers WH. Validation of a new three-item self-report measure for medication adherence. AIDS Behav 2016 Nov 20;20(11):2700-2708 [FREE Full text] [doi: 10.1007/s10461-016-1406-x] [Medline: 27098408]

28. Lauffenburger JC, Fontanet CP, Isaac T, Gopalakrishnan C, Sequist TD, Gagne JJ, et al. Comparison of a new 3-item self-reported measure of adherence to medication with pharmacy claims data in patients with cardiometabolic disease. Am Heart J 2020 Oct;228:36-43. [doi: 10.1016/j.ahj.2020.06.012] [Medline: 32768690]
29. Wolf MS, Smith SG, Pandit AU, Condon DM, Curtis LM, Griffith J, et al. Development and validation of the consumer health activation index. Med Decis Making 2018 Feb 13;38(3):334-343. [doi: 10.1177/0272989x17753392]
30. Gardner B, Abraham C, Lally P, de Bruijn G. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. Int J Behav Nutr Phys Act 2012 Aug 30;9(1):102 [FREE Full text] [doi: 10.1186/1479-5868-9-102] [Medline: 22935297]
31. Reis S, Biderman A, Mitki R, Borkan JM. Secrets in primary care: a qualitative exploration and conceptual model. J Gen Intern Med 2007 Sep 9;22(9):1246-1253 [FREE Full text] [doi: 10.1007/s11606-007-0186-4] [Medline: 17487521]
32. Borkan JM. Immersion/Crystallization. In: Crabtree BF, Miller WL, editors. Doing Qualitative Research. Thousand Oaks, California, United States: SAGE Publications; 1999:1-424.
33. Smith A. U.S. smartphone use in 2015. Pew Research Center. 2015. URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2021-05-15]
34. Kelly L, Jenkinson C, Morley D. Experiences of using web-based and mobile technologies to support self-management of type 2 diabetes: qualitative study. JMIR Diabetes 2018 May 11;3(2):e9 [FREE Full text] [doi: 10.2196/diabetes.9743] [Medline: 30291098]
35. Feldman J. The simplicity principle in perception and cognition. Wiley Interdiscip Rev Cogn Sci 2016 Sep 29;7(5):330-340 [FREE Full text] [doi: 10.1002/wcs.1406] [Medline: 27470193]
36. Baxter S, Enderby P, Evans P, Judge S. Barriers and facilitators to the use of high-technology augmentative and alternative communication devices: a systematic review and qualitative synthesis. Int J Lang Commun Disord 2012;47(2):115-129. [doi: 10.1111/j.1460-6984.2011.00090.x] [Medline: 22369053]
37. Ranney ML, Choo EK, Cunningham RM, Spirito A, Thorsen M, Mello MJ, et al. Acceptability, language, and structure of text message-based behavioral interventions for high-risk adolescent females: a qualitative study. J Adolesc Health 2014 Jul;55(1):33-40 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.12.017] [Medline: 24559973]
38. Moyano D, Morelli D, Santero M, Belizan M, Irazola V, Beratarrechea A. Perceptions and acceptability of text messaging for diabetes care in primary care in Argentina: exploratory study. JMIR Diabetes 2019 Mar 18;4(1):e10350 [FREE Full text] [doi: 10.2196/10350] [Medline: 30882362]
39. Bélanger E, Ahmed T, Vafaei A, Curcio CL, Phillips SP, Zunzunegui MV. Sources of social support associated with health and quality of life: a cross-sectional study among Canadian and Latin American older adults. BMJ Open 2016 Jun 28;6(6):e011503 [FREE Full text] [doi: 10.1136/bmjopen-2016-011503] [Medline: 27354077]
40. Litwin H. Social networks and well-being: a comparison of older people in Mediterranean and non-Mediterranean countries. J Gerontol B Psychol Sci Soc Sci 2010 Sep 14;65(5):599-608 [FREE Full text] [doi: 10.1093/geronb/gbp104] [Medline: 20008485]
41. Lawley KA, Willett ZZ, Scollon CN, Lehman BJ. Did you really need to ask? Cultural variation in emotional responses to providing solicited social support. PLoS One 2019 Jul 12;14(7):e0219478 [FREE Full text] [doi: 10.1371/journal.pone.0219478] [Medline: 31299054]
42. Oyserman D, Coon HM, Kemmelmeier M. Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. Psychol Bull 2002;128(1):3-72. [doi: 10.1037/0033-2909.128.1.3]
43. Torbjørnsen A, Ribu L, Rønnevig M, Grøttland A, Helseth S. Users' acceptability of a mobile application for persons with type 2 diabetes: a qualitative study. BMC Health Serv Res 2019 Sep 06;19(1):641 [FREE Full text] [doi: 10.1186/s12913-019-4486-2] [Medline: 31492176]
44. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
45. Rossmann C, Riesmeyer C, Brew-Sam N, Karnowski V, Joeckel S, Chib A, et al. Appropriation of mobile health for diabetes self-management: lessons from two qualitative studies. JMIR Diabetes 2019 Mar 29;4(1):e10271 [FREE Full text] [doi: 10.2196/10271] [Medline: 30924786]
46. Scheibe M, Reichelt J, Bellmann M, Kirch W. Acceptance factors of mobile apps for diabetes by patients aged 50 or older: a qualitative study. Med 2 0 2015 Mar 02;4(1):e1 [FREE Full text] [doi: 10.2196/med20.3912] [Medline: 25733033]
47. Baptista S, Wadley G, Bird D, Oldenburg B, Speight J, My Diabetes Coach Research Group. User experiences with a type 2 diabetes coaching app: qualitative study. JMIR Diabetes 2020 Jul 17;5(3):e16692 [FREE Full text] [doi: 10.2196/16692] [Medline: 32706649]
48. Keller PA, Lipkus IM, Rimer BK. Affect, framing, and persuasion. J Mark Res 2018 Oct 10;40(1):54-64. [doi: 10.1509/jmkr.40.1.54.19133]
49. Purnell JQ, Thompson T, Kreuter MW, McBride TD. Behavioral economics: "nudging" underserved populations to be screened for cancer. Prev Chronic Dis 2015 Jan 15;12:E06 [FREE Full text] [doi: 10.5888/pcd12.140346] [Medline: 25590600]
50. Gong J, Zhang Y, Yang Z, Huang Y, Feng J, Zhang W. The framing effect in medical decision-making: a review of the literature. Psychol Health Med 2013 Dec;18(6):645-653. [doi: 10.1080/13548506.2013.766352] [Medline: 23387993]

51. Piette JD, Farris KB, Newman S, An L, Sussman J, Singh S. The potential impact of intelligent systems for mobile health self-management support: Monte Carlo simulations of text message support for medication adherence. Ann Behav Med 2015 Feb 1;49(1):84-94 [FREE Full text] [doi: 10.1007/s12160-014-9634-7] [Medline: 25082177]
52. Hartz J, Yingling L, Powell-Wiley TM. Use of mobile health technology in the prevention and management of diabetes mellitus. Curr Cardiol Rep 2016 Dec 8;18(12):130. [doi: 10.1007/s11886-016-0796-8] [Medline: 27826901]
53. Levin IP, Schneider SL, Gaeth GJ. All frames are not created equal: a typology and critical analysis of framing effects. Organ Behav Hum Decis Process 1998 Nov;76(2):149-188. [doi: 10.1006/obhd.1998.2804] [Medline: 9831520]
54. Bullard KM, Cowie CC, Lessem SE, Saydah SH, Menke A, Geiss LS, et al. Prevalence of diagnosed diabetes in adults by diabetes type - United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Mar 30;67(12):359-361 [FREE Full text] [doi: 10.15585/mmwr.mm6712a2] [Medline: 29596402]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
EQUATOR: Enhancing the Quality and Transparency of Health Research
mHealth: mobile health

Edited by R Kukafka; submitted 22.11.20; peer-reviewed by M Pikkemaat, X Liao; comments to author 23.12.20; revised version received 19.01.21; accepted 30.04.21; published 11.06.21.

Please cite as:
Lauffenburger JC, Barlev RA, Sears ES, Keller PA, McDonnell ME, Yom-Tov E, Fontanet CP, Hanken K, Haff N, Choudhry NK
Preferences for mHealth Technology and Text Messaging Communication in Patients With Type 2 Diabetes: Qualitative Interview Study
J Med Internet Res 2021;23(6):e25958
URL: https://www.jmir.org/2021/6/e25958
doi:10.2196/25958
PMID:34114964

Original Paper

Co-designing a Lifestyle-Focused Text Message Intervention for Women After Breast Cancer Treatment: Mixed Methods Study

Anna Singleton[1,2], MSc, BSc(HonsI); Rebecca Raeside[1,2], MPH, BBiomedSc; Stephanie R Partridge[1,2,3], PhD, BSc; Molly Hayes[1,2], MBA, BCom; Katherine Maka[4,5], BAPSC; Karice K Hyun[1,2,6], PhD, MAS, BSc; Aravinda Thiagalingam[2,7,8], PhD, MBBS; Clara K Chow[2,7,8], PhD, MBBS; Kerry A Sherman[9], PhD, MA, BSc(Hons); Elisabeth Elder[5], PhD, MBBS; Julie Redfern[1,2,8,10], PhD, BAPSC

[1]Consumer Engagement and Codesign Research Hub, School of Health Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[2]Westmead Applied Research Centre, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[3]Prevention Research Collaboration, Charles Perkins Centre, University of Sydney, Sydney, Australia

[4]Department of Physiotherapy, Westmead Hospital, Sydney, Australia

[5]Westmead Breast Cancer Institute, Westmead Hospital, Sydney, Australia

[6]Department of Cardiology, Concord Repatriation General Hospital, Sydney, Australia

[7]Department of Cardiology, Westmead Hospital, Sydney, Australia

[8]George Institute for Global Health, University of New South Wales, Sydney, Australia

[9]Centre for Emotional Health, Department of Psychology, Macquarie University, Sydney, Australia

[10]Research Education Network, Western Sydney Local Health District, Sydney, Australia

Corresponding Author:
Anna Singleton, MSc, BSc(HonsI)
Consumer Engagement and Codesign Research Hub, School of Health Sciences
Faculty of Medicine and Health
University of Sydney
Level 6 Block K Westmead Hospital
Sydney, 2145
Australia
Phone: 61 288908181
Email: anna.singleton@sydney.edu.au

Abstract

Background: Breast cancer is the most common cancer among women globally. Recovery from breast cancer treatment can be mentally and physically challenging. SMS text message programs offer a novel way to provide health information and support, but few programs are co-designed with consumer representatives.

Objective: This study aims to report the procedures and outcomes of a co-design process of a lifestyle-focused SMS text message program to support women's mental and physical health after breast cancer treatment.

Methods: We followed an iterative mixed methods two-step process: (1) co-design workshop with consumers and health professionals and researchers to draft text messages and (2) evaluation of message content, which was scored (5-point Likert scale; 1=strongly disagree to 5=strongly agree) for ease of understanding, usefulness, and appropriateness, and readability (Flesch-Kincaid score). Additional free-text responses and semistructured interviews were coded into themes. Messages were edited or deleted based on the evaluations, with consumers' evaluations prioritized.

Results: In step 1, co-designed text messages (N=189) were semipersonalized, and the main content themes were (1) physical activity and healthy eating, (2) medications and side effects, (3) mental health, and (4) general breast cancer information. In step 2, consumers (n=14) and health professionals and researchers (n=14) provided 870 reviews of 189 messages and found that most messages were easy to understand (799/870, 91.8%), useful (746/870, 85.7%), and appropriate (732/870, 84.1%). However, consumers rated 50 messages differently from health professionals and researchers. On the basis of evaluations, 37.6% (71/189) of messages were deleted, 36.5% (69/189) were edited, and 12 new messages related to fatigue, self-care, and cognition were created. The final 130 text messages had a mean 7.12 (SD 2.8) Flesch-Kincaid grade level and 68.9 (SD 15.5) ease-of-reading score, which represents *standard* reading ease.

Conclusions: Co-designing and evaluating a bank of evidence-based mental and physical health-themed text messages with breast cancer survivors, health professionals, and researchers was feasible and resulted in a bank of 130 text messages evaluated highly by participants. Some consumer evaluations differed from health professionals and researchers, supporting the importance of co-design.

(J Med Internet Res 2021;23(6):e27076) doi:10.2196/27076

KEYWORDS

breast neoplasms; cancer survivors; text messaging; telemedicine; mobile health; co-design

Introduction

Background

Breast cancer is the most commonly diagnosed cancer among women globally [1]. More than 2 million women finish active breast cancer treatment (ie, surgery, chemotherapy, or radiotherapy) yearly [1], and many experience treatment-related mental and physical health challenges [2,3]. The 2013 Breast Health Global Initiative consensus statement posits that access to posttreatment information and support should be allocated to all patients [3]. However, there are limited survivorship care services [3]. Existing services usually occur in person, during work hours, or far from home, which increases the financial burden [4] and limits accessibility. Moreover, these services are rarely co-designed with patients [5,6]. Accessible posttreatment health support programs are required.

The Importance of Consumer Co-design

Co-design is the process by which service providers and consumers collaborate to develop meaningful and creative solutions. The benefits of co-design include improved services, provider-consumer interactions, and consumer engagement and experiences [7]. The key to effective co-design is making end-users' or consumers' lived experiences central throughout the project's lifespan [8] because it adds a unique and diverse perspective [8,9]. For example, *citizen collaborators* can be employed as active members of the research team from study conception to final dissemination [9]. Qualitative data suggest that access to posttreatment health information and support are among breast cancer survivors' top priorities for service improvement [2,10]. Enlisting breast cancer survivors as co-designers for novel posttreatment services may help place survivors' priorities at the forefront and strengthen the intervention's impact.

Co-design and Mobile Health Interventions

Mobile health (*mHealth*) uses mobile devices such as mobile phones to deliver health services [11]. mHealth interventions delivered by text messages are the most accessible to patients, as they are inexpensive, do not require internet for delivery, and can be sent anytime and anywhere with mobile services [12]. Moreover, text message programs effectively deliver health education to patients [12,13] and are easily co-designed [14]. Co-designed text message programs improved health-promoting behaviors among patients with heart disease [13] and were cost-saving for the health care system compared with traditional delivery methods [15]. A 2015 systematic review revealed that there were no text message interventions for cancer survivors [16]; more recently, there is some evidence that breast cancer survivors find the delivery of health support via text messages acceptable [17,18] and effective for improving adherence to endocrine therapy tablets [19,20] or maintain weight loss [21]. However, no studies reported that the interventions were co-designed and most focused on physical health outcomes, not including mental health messages within the intervention. Therefore, co-designing a text message intervention with women who completed active breast cancer treatment may be an innovative solution to reduce barriers to posttreatment health information and support.

Objectives

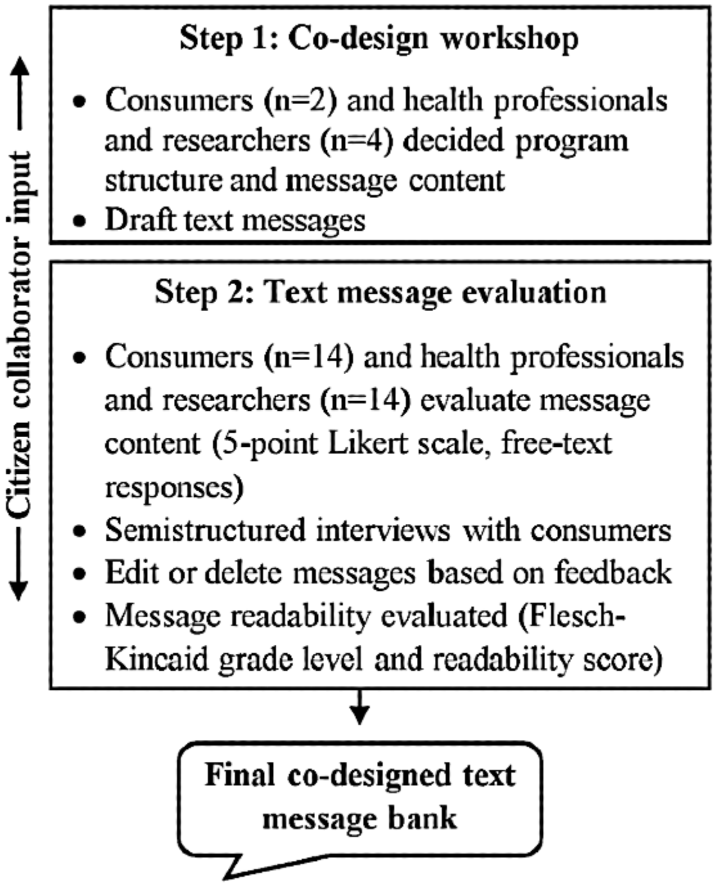
This study aims to report the procedures and outcomes of a co-design process of a bank of evidence-based text messages to support women's mental and physical health after breast cancer treatment. The co-designers were women who completed active breast cancer treatment (consumers and citizen collaborators), health professionals, and researchers. We hypothesize that consumers and citizen collaborators would add value by providing different inputs and evaluations regarding lived experience with breast cancer and the health care system compared with health professionals and researchers.

Methods

Study Design

A two-step mixed methods process [14] was used, including consumer representatives and a citizen collaborator (MH) as co-designers (Figure 1). Briefly, (step 1) an initial co-design workshop and (step 2) a structured qualitative text message evaluation and program refinement were conducted from June 2018 to March 2019 at the Westmead Breast Cancer Institute (WBCI), a major public breast cancer institute in New South Wales. WBCI resides within a large tertiary hospital in Western Sydney, Australia, serving a multicultural and diverse socioeconomic population (approximately 2.3 million people in 2016) [22]. The study was approved by the Western Sydney Local Health District Human Research Ethics Committee (AU RED HREC/18/WMEAD/281). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Written or electronic informed consent was obtained from all the participants.

Figure 1. Study design, including a citizen collaborator throughout.



Step 1: Co-design Workshop

The aim of step 1 was to work with consumer representatives, a citizen collaborator, health professionals, and researchers to develop the SMS text message program structure, content themes, and draft text message content.

Participants and Recruitment

Consumers were included if they had completed active breast cancer treatment (surgery, chemotherapy, or radiotherapy) within the past 5 years and did not have distant metastatic disease. Health professionals and researchers were eligible to participate if they were practicing health professionals (eg, medicine, allied health, or nursing) or active researchers with expertise in breast cancer or text message interventions. Consumers (n=3) from a local volunteer association, health professionals (n=8) from WBCI, and researchers (n=5) from 2 universities in Sydney, Australia, were invited via email to participate in a two-hour workshop to develop the program structure and message content. Those who agreed to attend were provided with further instructions regarding the workshop location. A citizen collaborator also reviewed the program structure and developed the message content. The citizen collaborator was aged 29 years, a White woman with lived experience with breast lumpectomy surgery, chemotherapy, radiotherapy, and 1 year of taking endocrine therapy tablets. The citizen collaborator assisted in developing the research idea and provided ongoing feedback throughout the program design.

Co-design Workshop

A two-hour interactive workshop was held on June 5, 2018. The workshop included a 15-minute presentation and semistructured discussion topics, such as the number of text messages delivered per week, if the delivery was one-way (no replies) or two-way (replies allowed), and message content themes. Time was also allocated to draft text message content. After the workshop, the health professionals drafted additional messages. The program structure and message content were discussed with the citizen collaborator before and after the workshop.

Theoretical Basis for Message Content

The text messages were developed based on the behavior change techniques taxonomy by Abraham and Michie [23], which includes the Bandura theory of social learning theory (social cognitive theory) and the information-motivation-behavioral skills model. The theories stipulate that providing information through education and motivation and observing others successfully complete behaviors can promote behavior change [24,25]. Moreover, assisting people in setting attainable goals and identifying barriers and strategies for overcoming barriers can facilitate goal achievement [24]. Table 1 provides examples of message content and corresponding behavior change techniques. In addition, researchers have found that people enjoy supportive personalized messages with a positive tone [26,27]. The messages were co-designed with these user satisfaction principles in mind, including achievable, practical tips and links to trustworthy websites and videos.

Table 1. Example of draft text message content and themes with corresponding behavior change techniques.

Behavior change technique	Example text message content	Message theme
Provide information about behavior-health link (IMB[a])	Fibre in fruit, veggies and wholegrain foods helps you feel full for longer, can improve blood sugars & lower cholesterol - so make sure there's plenty of fibre in your diet!	Nutrition
Provide information on consequences (TRA[b], TPB[c], SCogT[d], or IMB)	Research shows that consistent physical activity can help reduce antihormonal (endocrine) tablet side effects - so try to get moving every day but make sure to schedule time to rest!	Medication adherence and side effects
Prompt intention formation (TRA, TPB, SCogT, or IMB)	Sometimes we can do exercise without noticing - challenge yourself to park the car further away from the shops or your work so you get a few extra steps!	Physical activity
Prompt barrier identification (SCogT)	Side effects are the main reason why women stop taking their antihormonal (endocrine) tablets - Remember, [pref_name][e], your doctor has prescribed this medication to benefit your health and can help you if side effects are a problem.	Medication adherence or side effects
Provide general encouragement (SCogT)	Hi [pref_name], you are halfway through the EMPOWER-SMS[f] program – it's wonderful that you're taking a few moments each week to focus on your health and well-being - keep it up!	General
Set graded tasks (SCogT)	Sometimes getting started is the hardest part, [pref_name] - it can be easier to begin exercise at low intensity (walking, stretching) and gradually increase to moderate intensity (faster breathing rate but can still have a conversation)	Physical activity
Provide instruction (SCogT)	Physical activity is essential for your recovery - World Health Organization recommends at least 30mins of moderate-intensity activity 5days/week - click here for ideas: [insert link here]	Physical activity
Model or demonstrate the behavior (SCogT)	Some women experience tightness or restricted movement in their shoulder long after treatments have finished - if this is you, the BCI[g] shoulder care video may help: [insert link here]	Physical activity
Prompt specific goal setting (CT[h])	Hi [pref_name], relaxation and self-care are important - try making a list of 3 things that help you relax so you can include them into your week	Social or emotional well-being
Prompt self-monitoring of behavior	Do you have trouble sleeping? Consider keeping a sleep diary, it may help you understand what's impacting your sleep, like TV, lights or stress - see an example here: [insert link here]	General
Provide contingent rewards (OC[i])	Did you take your antihormonal (endocrine) tablet today, [pref_name]? If yes, well done! If not, don't worry, just take it as soon as you remember - if it's nearly time to take your next tablet, do not take two at the same time - 1 per day is enough	Medication adherence or side effects
Teach to use prompts or cues (OC)	Usually, people serve too much food on a large plate - reducing your plate size can help limit portion sizes and avoid overeating! - [centre_name]	Nutrition
Prompt practice (OC)	Writing a list of 3 things you're grateful for each week may change the way you feel – it's also a nice way to reflect - [centre_name]	Social or emotional well-being
Provide opportunities for social comparison (SCompT)	Hi [pref_name], having a hard time talking about your breast cancer? Breast Cancer Network Australia has a great web-based forum, where you can read women's questions and concerns, and you can pose questions, if you want. Here's the link: onlinenetwork.bcna.org.au	Social or emotional well-being
Plan social support or social change (social support theories)	Exercising with a friend can be fun & you can motivate each other - grab a friend and get moving or meet some new friends at a local walking group, find one near you here: [insert link here]	Physical activity
Prompt self-talk	Practicing positive self-talk is a good way to keep your mind healthy and improve your mood - not sure where to start? Click for more information: [insert link here]	Social or emotional well-being
Relapse prevention	Making small changes can make a big difference when it comes to healthy eating - at restaurants look for simple switches like a baked potato instead of chips. Read more tips here: [insert link here]	Nutrition
Stress management (stress theories)	Mindfulness meditation can help to manage stress and improve mood by focusing on breathing and relaxation techniques - if you'd like to learn more, click here: [insert link here]	Social or emotional well-being
Motivational interviewing	Remember to be kind to yourself, [pref_name]! Think of 3 things you're thankful that your body can do	General
Time management	Treat yourself to some *you* time, [pref_name] - whether it's a walk, yoga or reading a good book, setting aside time in your diary or on your phone can make it easier to stick to	Social or emotional well-being

[a] IMB: information-motivation-behavioral skills model.

[b] TRA: theory of reasoned action.

[c] TPB: theory of planned behavior.

[d] SCogT: social cognitive theory.

[e] [pref_name] is a place holder for the auto-population of the participant’s preferred name.

[f] EMPOWER-SMS: A text message intervention to support women’s mental and physical health after breast cancer treatment.

[g] BCI: breast cancer institute.

[h] CT: control theory.

[i] OC: operant conditioning.

Medical Content of Messages

Medical content was based on personal experiences of health professionals and confirmed using Cancer Australia’s national guidelines for primary prevention of cancer [28], managing menopausal symptoms after breast cancer treatment [29], and emerging breast cancer survivorship literature and systematic reviews.

The authors reviewed all messages (AS, JR, SRP, and RR) before text message evaluation and refinement (step 2). The reasons for deletions were recorded.

Step 2: Text Message Evaluation and Refinement

The aim of step 2 was for consumers, a citizen collaborator, health professionals, and researchers to evaluate the quality of the text message content, refine the messages based on feedback, and evaluate the messages’ ease of reading.

Participants and Recruitment

Consumers (n=16), the citizen collaborator, and health professionals and researchers (n=14) from the WBCI and 2 universities were invited to participate. Consumers were recruited during a medical clinic for one-year postsurgery follow-up appointments. Consumers were approached by their physician if they were an adult (>18 years) woman and completed active breast cancer treatment within the past 3 years (could still be taking endocrine therapy treatment) and were able to provide written informed consent. Consumers were not approached if they had metastatic breast cancer or insufficient English skills to provide informed consent or read and evaluate text messages.

Health professionals and researchers were eligible to participate if they were practicing health professionals (eg, medicine, allied health, or nursing) or active researchers with expertise in breast cancer or text message interventions. Health professionals and researchers known to the research team were invited to participate via email and provided electronic consent via a web-based database.

Text Message Evaluation by Consumers and Health Professionals

A combination of feedback surveys and semistructured interviews were conducted to evaluate the message content. For the feedback surveys, each participant provided basic demographic information regarding age, sex, and expertise (lived experience or health professional or researcher), and then evaluated 32 draft text messages (maximum 15 min) presented as 1 to 2 sentences on paper in a private room (consumers) or via a password-protected web-based database (health professionals or researchers). Owing to time constraints, one health professional evaluated only 16 messages. Message content related to medication adherence and side effects were reviewed by consumers who reported taking endocrine therapy tablets previously or at the time of the study.

The feedback survey contained 4 questions for each draft text message, with three 5-point Likert scale (1=strongly disagree to 5=strongly agree) questions: “This message was easy to understand;” “The information provided in this message is useful;” “This message is appropriate for women with breast cancer” and one free-text response question, “Do you have any suggestions to improve this message?” The Likert scale questions were summarized by the number and percentage of consumers’ or health professionals and researchers’ evaluations that *agreed* or *strongly agreed* that the messages were easy to understand, useful, and appropriate. Internal consistency was found among the 3 questions (Cronbach α=.88); therefore, each message received a mean score out of 15 (5 points per question; *strongly disagree*=1 point, *strongly agree*=5 points), which was the total score provided by each participant divided by the number of participants that evaluated the message. The mean consumer versus health professional and researcher score was also calculated for each message. Researchers were guided by a mean score <12 out of 15 (<80%), indicating less preference for a message from participants. All message scores and free-text replies were reviewed in parallel to guide message edits or deletions. If consumers and health professionals or researchers’ ratings differed, priority was given to consumers’ evaluations.

Free-text consumer and health professional or researcher evaluations were collated for each message. The citizen collaborator evaluated all messages via free-text feedback. After completing the feedback survey, consumers were invited to brief (10-15 min) semistructured interviews that explored participants’ most and least liked text messages, suggestions for additional message themes, the appropriateness of the language within the messages (too complicated vs too simple), and whether a SMS text message program would have been helpful during the transition from hospital care to health self-management. The interviews were audio-recorded and transcribed verbatim. The free-text feedback and interview transcripts were open-coded using inductive thematic content analysis in NVivo 11 (QSR International) by 2 independent parallel coders (AS and RR). The researchers discussed codes until they reached an agreement on themes. The feedback survey results and interviews were used to inform message refinement (edits, deletions, or creation of new message content).

Evaluation of Readability

All messages were evaluated for readability using the Flesch-Kincaid ease-of-reading score and grade level, which are validated scores of content readability based on the number of words per sentence and syllables per word [30,31]. The Flesch readability score ranges from 0 to 100, with 0 being extremely difficult and 100 being extremely easy to read. Grade levels range from 1 to 12; grade 1 is the easiest, and grade 12 is the hardest to read. The characters and word counts of the text messages were also calculated.

Results

Step 1: Co-design Workshop

Consumers (n=2) and health professionals and researchers (n=4; health psychologist, specialist breast cancer surgeon, physiotherapist and digital health researcher, and psychology researcher) attended the workshop. The group decided that the program would be one-way (no replies) as consumers did not want to feel pressure to reply. Our team's previous research also showed that one-way messaging was well-liked by patients with heart disease and helped them feel supported [27]. Next, the group decided that the text messages would be delivered randomly 4 times per week, as consumers felt that 1-2 messages were too few to be useful but 6 or 7 would be annoying. In our previous research, 4 messages per week were found to be acceptable [27]. Unlike our team's previous research, which only delivered messages on weekdays [14], the consumers desired messages to be delivered Monday-Saturday (maximum 1 message per day) because people may have more time to read messages on the weekend. Messages would be delivered at 9 AM, noon, 3 PM, or 6 PM to mimic the unpredictability of how a friend or family member would text and gain insights into future process evaluations as to what times are best. The team decided that 1 message per theme would be delivered per week in random order, as consumers indicated that a specific order was not required. The co-designed message content themes were (1) physical activity and healthy eating, (2) endocrine therapy tablet (medication) adherence and side effects, (3) social and emotional well-being, and (4) general breast cancer information. Moreover, participants wanted messages to contain links to trustworthy websites because it can be challenging to navigate cancer information on the web. It was also agreed that the program would be semipersonalized according to the participants' preferred name or *nickname*, and there would be specific messages for participants taking endocrine therapy tablets because these tablets can cause challenging side effects and medication nonadherence [32].

A total of 274 messages were co-designed; 157 messages were drafted during the workshop, and 117 were created by the authors, including the citizen collaborator, after the workshop. The citizen collaborator's content subtheme ideas included posttreatment financial support, difficulties returning to work (fatigue, treatment side effects, or cognitive load), and practical solutions for medication side effects. After revision of the messages by authors AS, JR, SRP, and RR, 189 co-designed messages remained for evaluation in step 2. Reasons for deletion included repetitive ideas or content (65/274, 24%), negative message tone (11/274, 4%), and message content not being relevant to breast cancer survivors (9/274, 4%)

Step 2: Text Message Content Evaluation

Participants

Consumers (14/16, 87.5%) and health professionals or researchers (14/14, 100%) completed the message evaluation; 2 consumers (2/16, 12.5%) did not return their surveys and were not included in the analyses. The mean age of consumers was 60 (SD 10) years, and the mean age of health professionals or researchers was 46 (SD 8) years. Most participants were White (12/14, 86% of consumers; 12/14, 86% of health professionals or researchers) women (14/14, 100% of consumers; 12/14, 86% of health professionals or researchers).

Consumer and Health Professional Evaluation

The 189 co-designed text messages were evaluated 870 times: 443 times by consumers, and 427 times by health professionals or researchers (mean 2.34 consumer and 2.26 health professional evaluations per message). Overall, participants *agreed* or *strongly agreed* that the messages were clear (799/870, 91.8%), useful (746/870, 85.7%), and appropriate (732/870, 84.1%). Most messages (156/189, 82.5%) received a mean score of ≥12 out of 15 from consumers (153/189, 80.9%) and health professionals (156/189, 82.5%). However, 50 messages received a different mean score from consumers versus health professionals or researchers; 25 were evaluated highly by consumers (mean score ≥12/15, 80%) but not health professionals or researchers, and the other 25 were evaluated poorly by consumers (mean score <12/15, 80%). Examples of discrepant evaluation scores and free-text feedback are presented in Table 2. The number of free-text responses was similar between health professionals (n=150) and consumers (n=156), suggesting that the media on which they conducted the surveys (in person vs on the web) did not bias the feedback. Free-text feedback themes from health professionals or researchers (n=14) focused on grammar, technical terms, and adding web links. Consumers' (n=14) free-text feedback focused on personal experiences and increased positivity in messages.

Table 2. Example text messages and decisions based on discrepant consumers' versus health professionals and researchers' mean ratings (out of 15) and supporting feedback quotes.

Message themes and subthemes		Decision
General breast cancer information		
Returning to work		
"Hi [preferred name], returning to work can be difficult - some women have found it helpful to openly discussing their needs and working arrangements with their managers so that everyone is on the same page"	• Consumers' mean rating 10/15 (67%): "Work were not helpful and sick leave all used up - refused flexible work and had to feel guilty. Head of [Human Resources] however was supportive, not line manager and above" (female, age 65 years) • Health professional and researchers' mean rating 12/15 (80%): "wording change after managers 'discussing' to discuss eg: 'so a plan can be made' or just stop after managers" (female, age 62 years)	Deleted
Link to web-based breast cancer fact sheets or resources		
"Did you know that the Breast Cancer Network Australia provides a variety of kits, booklets and fact sheets where you can find more information? Check it out: [insert link]"	• Consumers' mean rating 13.5/15 (90%): "helpful" (female, age 65 years) • Health professionals and researchers' mean rating 11/15 (73%): "All patients with breast cancer get this information at the time of their diagnosis" (female, age 54 years)	Edited
Medication adherence and side effects		
Hot flushes		
"If you are experiencing hot flushes at night, give a cool shower right before bed a go and see if that helps!"	• Consumers' mean rating 12/15 (80%): "aware this is a side effect especially of hormonal treatment" (female, age 65 years) • Health professionals and researchers' mean rating 10/15 (67%): "'Have a cold shower' I'm not sure that cold showers will be popular??" (Male, age 51 years); "'Wording not good – suggest'...taking a cool shower just before going to bed might help' or something similar" (female, age 54 years)	Deleted
Physical activity		
Counting steps		
"Apps, watches or other step counters can help with tracking your steps each day - The World Health Organisation recommends 10,000 steps per day. How many do you usually do [preferred name]?"	• Consumers' mean rating 11.5/15 (77%): "While the message is clear, I don't think I need to receive it. Messages like these are not personal and supportive. Just that not everyone has apps, younger generation like my daughter does, with her fitbit she tracks steps. I don't have one. A message like 30 mins (walking) of exercise 3-5x a week is recommended, how often do you exercise" (female, age 45 years) • Health professionals and researchers' mean rating 15/15 (100%): "should there be a link [or] reasoning why useful for [Breast Cancer Institute] patient" (female, age 38 years)	Deleted. Consumer's suggestion is referring to another message she evaluated, which is included in the final message bank.
Nutrition		
Portion sizes		
"Hey [preferred name], some people eat too much of a good thing. As a general rule - it's good to keep portion sizes around the size of your fist"	• Consumers' mean rating 13.5/15 (90%): "Good information" (female, age 50 years) • Health professional and researchers' mean rating 11.5/15 (77%): "Could delete. Fruit and veg portions is larger than fist" (female, age 31 years)	Edited
Social and emotional well-being		
Setting goals for self-care		
"Hi [preferred name], setting aside time for yourself in your diary or on your phone can make it easier to stick to - whether it's a walk, yoga or just reading a good book, treat yourself to some 'you' time"	• Consumers' mean rating 12.33/15 (82%): "resources need to be more available. [they are] not always in a prominent place. These resources would have helped - had to really search." (female, age 65 years) • Health professionals and researchers' mean rating 10/15 (67%): "This message seems too busy. I would simplify. Maybe make into 2 messages as you are tapping into 2 different ideas: regular practice and 'me time'. I would separate." (female, age 55 years)	Edited

Oral interview feedback from consumers (n=3) and free-text feedback from the citizen collaborator revealed three important themes for consideration in message refinement: prevention of loneliness, impacts of treatment and medication side effects, and motivation to exercise. Example quotes for each theme are presented in Textbox 1. Feedback also revealed that the term *journey* was polarizing. For example, a consumer stated, “Many women don’t like the word journey. Could use breast cancer experiences instead” (female, age 41 years), and a health professional’s feedback mirrored this sentiment: “some people don’t like the word journey, maybe consider alternative eg experience” (female, age 62 years). Therefore, *journey* was replaced with the term *experience* throughout. Consumers also felt that messages relating to alcohol and cooking food on a barbeque were not appropriate in some cultures. For example, a consumer stated, “I heard the BBQ is not good for cancer people because when protein and oil getting burnt...burnt food is no good for cancer” and later explained, “here is one about limit[ing] alcohol. It’s a good message but some people when they read that message, even they don’t want to drink alcohol. Some people read it and they might think ‘oh maybe I can drink alcohol’” (consumer, female, age 50 years). All messages about alcohol were deleted, and the word *BBQ* was replaced with a suggestion to *bake* the food.

Textbox 1. Consumer and citizen collaborator message content themes and quotes.

Theme 1: Prevention of Loneliness and Isolation

> *The messages [that] give information, give some encouragement because after treatment, I stayed home and I felt very lonely.* [Consumer, female, age 50 years]

Theme 2: Impact of Side Effects

> *I was slowly recuperating pretty good, but I’m still not over it yet, it’s been over 12 months...Tiredness, sometimes I can’t do housework.* [Consumer, female, age 76 years]
>
> *Fatigue was huge. And it would always hit you sort of 2, 3 o’clock in the afternoon.* [Consumer, female, age 56 years]

Theme 3: Motivation for Exercise

> *have an exercise buddy or just a friend that you meet to go and have a walk with or walk your dog...just go I’m meeting so and so and we’re going to do this today, and it’s like keeping an appointment* [Consumer, female, age 56 years]
>
> *[My smartwatch] counted the amount of steps I did every day...so it would tell me how many kilometres I had done. It gives you this sense of achievement at the end of the week.* [Consumer, female, age 56 years]
>
> *One thing you could add is a reminder that finding time to exercise and eat healthy can be really hard when you struggle with fatigue, but that it will actually help give you more energy.* [Citizen collaborator, female, age 30 years]
>
> *after operation...it’s been 3 years...but still pain and the scar is hard. When I’m moving it feels hard [tight]. The doctor said more exercises, stretch and do some physical therapy.* [Consumer, female, age 50 years]

On the basis of the feedback (Likert scale, free-text questions, and interviews), 36.5% (69/189) of messages were edited, 37.6% (71/189) were deleted (score <12 or repetitive concept), and 12 new messages were created; 130 evidence-based, co-designed messages remained: 55.4% (72/130) of messages relating to the theme of physical activity and healthy eating, 29.2% (38/130) regarding social and emotional well-being, 32.3% (42/130) regarding general breast cancer information, and 28.5% (37/130) regarding medication adherence and side effects. Subthemes from step 1, such as financial support and practical solutions for side effects (fatigue, hot flushes, and cognitive load), were included within these four broader themes.

Readability

The 130 text messages had a mean Flesch-Kincaid grade level of 7.12 (SD 2.84) and an ease-of-reading score of 68.92 (SD 15.46), representing *standard* reading ease, which means that the texts will be easy to read for anyone with grade (or year) 7 education and higher [31]. The text messages had a mean of 2.02 (SD 0.69) sentences, 29.09 (SD 6.72) words, or 132.74 (SD 31.69) characters.

Discussion

Principal Findings

This study reports the outcomes of a co-design process [14] and the development of a bank of evidence-based text messages. To our knowledge, this is the first SMS text message intervention to be co-designed by breast cancer survivors, a citizen collaborator, health professionals, and researchers, and includes both mental and physical health messages. The message content was combined with clinical guidelines [28] and behavior change techniques [23]. Messages were evaluated highly for ease of understanding, usefulness, and appropriateness for breast cancer survivors by consumers, health professionals, and researchers. Although there were discrepancies between some consumers’ and health professionals and researchers’ mean message ratings, the consumers’ feedback was prioritized to ensure that their voices were at the forefront of the co-design process. Compared with the initial bank of draft text messages (n=274), the resulting 130 evidence-based co-designed text messages were evaluated highly (score ≥12/15, 80% and positive qualitative feedback) by participants and focused on topics of interest to consumers—namely, social and emotional well-being, healthy eating, physical activity, medication adherence, side effects, and general breast cancer information. Overall, this iterative co-design process was feasible, produced an improved final text message bank, and highlighted the importance of consumer co-design.

SMS text message interventions are becoming increasingly popular as mHealth strategies [33]. However, there is sparse research on women with breast cancer [16-21], and few programs describe the process of text message content development [34]. For example, health literacy is often overlooked or not reported in the development of eHealth programs [35]. In this study, messages were short and rated highly for ease of understanding, which is supported by the Flesch-Kincaid grade level 7 and readability score (SD 68.9). These scores indicate that anyone with an education grade of 7 or above would likely understand the messages [31], which may reduce health literacy barriers [30] and increase user acceptability [27]. Moreover, health professionals, researchers, and consumers feedback revealed that a positive message tone was important, which is consistent across studies [27,36]. Recent systematic reviews reveal that message tailoring and personalization may [37] or may not [38] improve program efficacy, and message content does not need to be based on a theoretical framework to be effective [38]. Qualitative results suggest that it may be the overall impact of the program rather than a single program component that end users find helpful [27]. However, for adequate comparison across interventions, the development of message content should be clearly reported [37,38]. Factors including the co-design strategy, behavior change theories, readability, and message personalization can help elucidate what makes an effective and well-liked intervention.

This study also provides a framework for developing and implementing SMS text message programs with consumers' voices at the forefront. Previous studies have developed and refined messages with health professionals, and consumers reviewed the edited messages [14,39]. This strategy may miss message topics or lived experiences important to consumers and limit program and delivery impact. Conversely, this study simultaneously involved consumers and a citizen collaborator with health professionals and researchers throughout the message development, review, and refinement. Moreover, all participants (consumers, a citizen collaborator, health professionals, and researchers) reviewed the same initial messages and were provided equal opportunity for opinions; furthermore, we prioritized consumer ratings when refining messages. Research shows that developing programs with consumers and citizen collaborators can improve user acceptance and engagement [7,9], which may explain the high rating of message usefulness and acceptability from both consumers and health professionals and researchers in this study. However, an ongoing challenge is to provide adequate compensation to consumers for their involvement and efforts. The National Health and Medical Research Council recently released guidelines for the involvement of consumers in research, which suggest that consumer reimbursement (financial, authorship, or acknowledgment) should be considered when planning a project or applying for grants [40]. These guidelines may help future researchers to identify barriers and enablers in consumer co-design.

Implications for Practice

The proliferation of breast cancer survivors is a testament to improvement in treatment [1]. However, women's mental health can be impacted years after a breast cancer diagnosis [41,42] and increase the urgency for ongoing supportive care. There is evidence that eHealth self-management programs *during* cancer treatment can reduce women's distress and depression [43,44]. However, few programs address the mental health concerns that arise *after* treatment [45]. In fact, many text message programs focus on improving physical behaviors such as physical activity, healthy eating, smoking cessation, and medication adherence [34]. However, previous research shows that mental and physical health are closely related and that consumers desire mental health information [5,46,47]. This program endeavored to fill this gap by co-designing messages for mental health after breast cancer treatment. The consumer feedback from this study provided further evidence that loneliness and treatment side effects are ongoing concerns that may impact mental health. This world-first co-designed program for breast cancer survivors may therefore add to current posttreatment support services by offering an overall lifestyle-focused approach, including both mental and physical health promotion.

Although the study provides new evidence for the co-design of SMS text message interventions for breast cancer survivors, it has some limitations. The number of consumers was relatively small, and most were White, which may not be generalizable to the priorities of the wider breast cancer community. However, cultural considerations, such as alcohol intake and cooking styles, were considered in the message evaluations and refinements. Consumers were recruited from WBCI, which services a large (about 2.3 million) multilingual and socioeconomically diverse population in Western Sydney. Half (49%) of the Western Sydney population speak a non-English language at home, including 8.5% who reported speaking English *not well* or *not at all* [22]. It is possible that the requirement to read and understand English limited some people from participating. However, this study demonstrates that the co-design process is acceptable and can be easily replicated in other populations or languages. To increase the diversity and number of message reviewers, future research should provide videoconference options for co-design workshops and web-based options for text message evaluation surveys, where possible and allowable through ethics. Reducing access barriers will allow women from distant or rural communities to offer important inputs. Moreover, the final text message bank will be evaluated in a randomized controlled trial and process evaluation [48]. A diverse sample of 160 women will be recruited to provide feedback via a questionnaire and focus groups to elucidate the feasibility and acceptability of the message content and overall one-way program within the wider breast cancer community.

Conclusions

Consumers, a citizen collaborator, health professionals, and researchers successfully co-designed a set of 130 evidence-based lifestyle-focused text messages to support women's health after breast cancer treatment. The co-design process resulted in an improved final bank of text messages and highlighted the importance of involving consumers as equal co-designers. This process can easily be replicated to develop SMS text message interventions for other patient populations.

Acknowledgments

The authors thank the breast cancer consumer representatives from the Breast Cancer Network Australia and WBCI and all staff and patients at the WBCI for their involvement in the study. A special thanks to the consumer representatives Christine Mitchell and Fiona Neill. Thanks to the Westmead Applied Research Center staff members for their research support, especially Caroline Wu. AS had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

No project funding was received for this study, but several team members were supported by scholarships and fellowships. AS is supported by the University of Sydney's Research Training Program Scholarship and the Westmead Applied Research Center Supplementary Postgraduate Research Scholarship in Breast Cancer. SRP is supported by a National Health and Medical Research Council Early Career Fellowship **(**APP1157438) and National Heart Foundation Postdoctoral Fellowship (HF102164). KKH is supported by a National Heart Foundation Postdoctoral Fellowship (HF 102138). JR is supported by the National Health and Medical Research Council Career Development Fellowship (APP1143538). CKC is supported by a Career Development Fellowship level 2 cofunded by the National Health and Medical Research Council and National Heart Foundation Future Leader Award (APP1105447). The data that support the findings of this study are available from the corresponding author (AS) upon reasonable request.

Authors' Contributions

AS, JR, SRP, EE, and MH contributed to the study conception and design. AS and RR supported the details of recruitment, data collection, and analyses of the results. AS, JR, SRP, and RR drafted the manuscript. All authors supported the text message development, intervention design, commented on previous versions of the manuscript and have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Fiszer C, Dolbeault S, Sultan S, Brédart A. Prevalence, intensity, and predictors of the supportive care needs of women diagnosed with breast cancer: a systematic review. Psychooncology 2014 Apr;23(4):361-374. [doi: 10.1002/pon.3432] [Medline: 24677334]
3. Ganz P, Yip C, Gralow J, Distelhorst S, Albain K, Andersen B, et al. Supportive care after curative treatment for breast cancer (survivorship care): resource allocations in low- and middle-income countries. A Breast Health Global Initiative 2013 consensus statement. Breast 2013 Oct;22(5):606-615. [doi: 10.1016/j.breast.2013.07.049] [Medline: 24007941]
4. Miedema B, Easley J. Barriers to rehabilitative care for young breast cancer survivors: a qualitative understanding. Support Care Cancer 2012 Jun;20(6):1193-1201. [doi: 10.1007/s00520-011-1196-7] [Medline: 21625913]
5. Mead KH, Raskin S, Willis A, Arem H, Murtaza S, Charney L, et al. Identifying patients' priorities for quality survivorship: conceptualizing a patient-centered approach to survivorship care. J Cancer Surviv 2020 Dec;14(6):939-958. [doi: 10.1007/s11764-020-00905-8] [Medline: 32607715]
6. Nekhlyudov L, Mollica M, Jacobsen P, Mayer D, Shulman L, Geiger A. Developing a Quality of Cancer Survivorship Care Framework: Implications for Clinical Care, Research, and Policy. J Natl Cancer Inst 2019 Nov 01;111(11):1120-1130 [FREE Full text] [doi: 10.1093/jnci/djz089] [Medline: 31095326]
7. Steen M, Manschot M, De KN. Benefits of co-design in service design projects. Int J Design 2011;5(2) [FREE Full text]
8. Tsianakas V, Robert G, Maben J, Richardson A, Dale C, Griffin M, et al. Implementing patient-centred cancer care: using experience-based co-design to improve patient experience in breast and lung cancer services. Support Care Cancer 2012 Nov;20(11):2639-2647 [FREE Full text] [doi: 10.1007/s00520-012-1470-3] [Medline: 22544223]
9. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect 2014 Oct;17(5):637-650 [FREE Full text] [doi: 10.1111/j.1369-7625.2012.00795.x] [Medline: 22809132]
10. Tsianakas V, Maben J, Wiseman T, Robert G, Richardson A, Madden P, et al. Using patients' experiences to identify priorities for quality improvement in breast cancer care: patient narratives, surveys or both? BMC Health Serv Res 2012 Aug 22;12:271 [FREE Full text] [doi: 10.1186/1472-6963-12-271] [Medline: 22913525]

11. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
12. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
13. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of Lifestyle-Focused Text Messaging on Risk Factor Modification in Patients With Coronary Heart Disease: A Randomized Clinical Trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
14. Redfern J, Thiagalingam A, Jan S, Whittaker R, Hackett ML, Mooney J, et al. Development of a set of mobile phone text messages designed for prevention of recurrent cardiovascular events. Eur J Prev Cardiol 2014 Apr;21(4):492-499. [doi: 10.1177/2047487312449416] [Medline: 22605787]
15. Burn E, Nghiem S, Jan S, Redfern J, Rodgers A, Thiagalingam A, et al. Cost-effectiveness of a text message programme for the prevention of recurrent cardiovascular events. Heart 2017 Jun;103(12):893-894. [doi: 10.1136/heartjnl-2016-310195] [Medline: 28235776]
16. Goode AD, Lawler SP, Brakenridge CL, Reeves MM, Eakin EG. Telephone, print, and Web-based interventions for physical activity, diet, and weight control among cancer survivors: a systematic review. J Cancer Surviv 2015 Dec;9(4):660-682. [doi: 10.1007/s11764-015-0442-2] [Medline: 25757733]
17. M Quintiliani L, Mann DM, Puputti M, Quinn E, Bowen DJ. Pilot and Feasibility Test of a Mobile Health-Supported Behavioral Counseling Intervention for Weight Management Among Breast Cancer Survivors. JMIR Cancer 2016 May 09;2(1):e4 [FREE Full text] [doi: 10.2196/cancer.5305] [Medline: 28410174]
18. Job JR, Spark LC, Fjeldsoe BS, Eakin EG, Reeves MM. Women's Perceptions of Participation in an Extended Contact Text Message-Based Weight Loss Intervention: An Explorative Study. JMIR Mhealth Uhealth 2017 Feb 27;5(2):e21 [FREE Full text] [doi: 10.2196/mhealth.6325] [Medline: 28242595]
19. Mougalian SS, Epstein LN, Jhaveri AP, Han G, Abu-Khalaf M, Hofstatter EW, et al. Bidirectional text messaging to monitor endocrine therapy adherence and patient-reported outcomes in breast cancer. JCO Clin Cancer Inform 2017 Nov;1:1-10 [FREE Full text] [doi: 10.1200/CCI.17.00015] [Medline: 30657377]
20. Hershman DL, Unger JM, Hillyer GC, Moseley A, Arnold KB, Dakhil SR, et al. Randomized Trial of Text Messaging to Reduce Early Discontinuation of Adjuvant Aromatase Inhibitor Therapy in Women With Early-Stage Breast Cancer: SWOG S1105. J Clin Oncol 2020 Jul 01;38(19):2122-2129. [doi: 10.1200/JCO.19.02699] [Medline: 32369401]
21. Spark LC, Fjeldsoe BS, Eakin EG, Reeves MM. Efficacy of a Text Message-Delivered Extended Contact Intervention on Maintenance of Weight Loss, Physical Activity, and Dietary Behavior Change. JMIR Mhealth Uhealth 2015 Sep 15;3(3):e88 [FREE Full text] [doi: 10.2196/mhealth.4114] [Medline: 26373696]
22. Hamilton O. Cultural and Linguistic Diversity in Greater Western Sydney. WESTIR Ltd. 2016. URL: https://www.westir.org.au/new/images/CALDOCC.pdf [accessed 2021-04-07]
23. 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]
24. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
25. Strecher VJ, DeVellis BM, Becker MH, Rosenstock IM. The role of self-efficacy in achieving health behavior change. Health Educ Q 1986;13(1):73-92. [doi: 10.1177/109019818601300108] [Medline: 3957687]
26. Bock BC, Heron KE, Jennings EG, Magee JC, Morrow KM. User preferences for a text message-based smoking cessation intervention. Health Educ Behav 2013 Apr;40(2):152-159. [doi: 10.1177/1090198112463020] [Medline: 23086555]
27. Redfern J, Santo K, Coorey G, Thakkar J, Hackett M, Thiagalingam A, et al. Factors influencing engagement, perceived usefulness and behavioral mechanisms associated with a text message support program. PLoS One 2016 Oct 14;11(10):e0163929 [FREE Full text] [doi: 10.1371/journal.pone.0163929] [Medline: 27741244]
28. Position Statement - Lifestyle Risk Factors and the Primary Prevention of Cancer. Australian Government - Cancer. 2015. URL: https://www.canceraustralia.gov.au/publications-and-resources/position-statements/lifestyle-risk-factors-and-primary-prevention-cancer [accessed 2021-05-21]
29. Cancer Australia. Managing Menopausal Symptoms After Breast Cancer - A Guide for Women Second Edition. Cancer Australia Publications. 2016. URL: https://tinyurl.com/4h63p4d6 [accessed 2021-05-21]
30. Williamson J, Martin A. Analysis of patient information leaflets provided by a district general hospital by the Flesch and Flesch-Kincaid method. Int J Clin Pract 2010 Dec;64(13):1824-1831. [doi: 10.1111/j.1742-1241.2010.02408.x] [Medline: 21070533]
31. Flesch R. A new readability yardstick. J Appl Psychol 1948 Jun;32(3):221-233. [doi: 10.1037/h0057532] [Medline: 18867058]
32. Brett J, Fenlon D, Boulton M, Hulbert-Williams N, Walter F, Donnelly P, et al. Factors associated with intentional and unintentional non-adherence to adjuvant endocrine therapy following breast cancer. Eur J Cancer Care (Engl) 2018 Jan;27(1):e12601 [FREE Full text] [doi: 10.1111/ecc.12601] [Medline: 27901302]

33. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct 17;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
34. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
35. 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]
36. Ranney M, Choo E, Cunningham R, Spirito A, Thorsen M, Mello M, et al. Acceptability, language, and structure of text message-based behavioral interventions for high-risk adolescent females: a qualitative study. J Adolesc Health 2014 Jul;55(1):33-40 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.12.017] [Medline: 24559973]
37. Head K, Noar S, Iannarino N, Harrington NG. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48 [FREE Full text] [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
38. Armanasco A, Miller Y, Fjeldsoe B, Marshall A. Preventive Health Behavior Change Text Message Interventions: A Meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
39. Fletcher R, May C, Wroe J, Hall P, Cooke D, Rawlinson C, et al. Development of a set of mobile phone text messages designed for new fathers. J Reproduct Infant Psychol 2016 Aug 9;34(5):525-534. [doi: 10.1080/02646838.2016.1214250]
40. Statement on Consumer and Community Involvement in Health and Medical Research. National Health and Medical Research Centre. 2016. URL: https://www.nhmrc.gov.au/about-us/publications/statement-consumer-and-community-involvement-health-and-medical-research [accessed 2021-04-05]
41. Maass S, Roorda C, Berendsen A, Verhaak P, de Bock GH. The prevalence of long-term symptoms of depression and anxiety after breast cancer treatment: A systematic review. Maturitas 2015 Sep;82(1):100-108 [FREE Full text] [doi: 10.1016/j.maturitas.2015.04.010] [Medline: 25998574]
42. Zainal NZ, Nik-Jaafar NR, Baharudin A, Sabki ZA, Ng CG. Prevalence of depression in breast cancer survivors: a systematic review of observational studies. Asian Pac J Cancer Prev 2013 Apr 30;14(4):2649-2656 [FREE Full text] [doi: 10.7314/apjcp.2013.14.4.2649] [Medline: 23725190]
43. Kim SC, Shaw BR, Shah DV, Hawkins RP, Pingree S, McTavish FM, et al. Interactivity, Presence, and Targeted Patient Care: Mapping e-Health Intervention Effects Over Time for Cancer Patients with Depression. Health Commun 2019 Feb;34(2):162-171 [FREE Full text] [doi: 10.1080/10410236.2017.1399504] [Medline: 29135321]
44. Lally RM, Kupzyk KA, Bellavia G, Hydeman J, Gallo S, Helgeson VS, et al. CaringGuidance™ after breast cancer diagnosis eHealth psychoeducational intervention to reduce early post-diagnosis distress. Support Care Cancer 2020 May;28(5):2163-2174 [FREE Full text] [doi: 10.1007/s00520-019-05028-0] [Medline: 31414245]
45. Willems RA, Mesters I, Lechner L, Kanera IM, Bolman CA. Long-term effectiveness and moderators of a web-based tailored intervention for cancer survivors on social and emotional functioning, depression, and fatigue: randomized controlled trial. J Cancer Surviv 2017 Dec;11(6):691-703 [FREE Full text] [doi: 10.1007/s11764-017-0625-0] [Medline: 28698999]
46. Thakkar J, Redfern J, Thiagalingam A, Chow CK. Patterns, predictors and effects of texting intervention on physical activity in CHD - insights from the TEXT ME randomized clinical trial. Eur J Prev Cardiol 2016 Nov;23(17):1894-1902. [doi: 10.1177/2047487316664190] [Medline: 27512051]
47. Ohrnberger J, Fichera E, Sutton M. The relationship between physical and mental health: A mediation analysis. Soc Sci Med 2017 Dec;195:42-49 [FREE Full text] [doi: 10.1016/j.socscimed.2017.11.008] [Medline: 29132081]
48. Singleton A, Partridge SR, Raeside R, Regimbal M, Hyun KK, Chow CK, et al. A text message intervention to support women's physical and mental health after breast cancer treatments (EMPOWER-SMS): a randomised controlled trial protocol. BMC Cancer 2019 Jul 04;19(1):660. [doi: 10.1186/s12885-019-5886-8] [Medline: 31272399]

Abbreviations

mHealth: mobile health
WBCI: Westmead Breast Cancer Institute

Edited by R Kukafka; submitted 10.01.21; peer-reviewed by M Bientzle, A Bermon, J Job, R Brar Prayaga; comments to author 17.03.21; revised version received 07.04.21; accepted 06.05.21; published 14.06.21.

Please cite as:
Singleton A, Raeside R, Partridge SR, Hayes M, Maka K, Hyun KK, Thiagalingam A, Chow CK, Sherman KA, Elder E, Redfern J
Co-designing a Lifestyle-Focused Text Message Intervention for Women After Breast Cancer Treatment: Mixed Methods Study
J Med Internet Res 2021;23(6):e27076
URL: https://www.jmir.org/2021/6/e27076
doi:10.2196/27076
PMID:34125072

Original Paper

A Comprehensive 6A Framework for Improving Patient Self-Management of Hypertension Using mHealth Services: Qualitative Thematic Analysis

Ting Song[1,2], MSc; Fang Liu[3], MD; Ning Deng[4], PhD; Siyu Qian[1,2,5], PhD; Tingru Cui[6], PhD; Yingping Guan[3], BSc; Leonard Arnolda[2], PhD; Zhenyu Zhang[1], MSc; Ping Yu[1,2], PhD

[1]Centre for Digital Transformation, School of Computing and Information Technology, Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, Australia

[2]Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, Australia

[3]Department of Health Examination, General Hospital of Ningxia Medical University, Yinchuan, China

[4]Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

[5]Drug and Alcohol Service, Illawarra Shoalhaven Local Health District, Wollongong, Australia

[6]School of Computing and Information Systems, Faculty of Engineering and Information Technology, University of Melbourne, Melbourne, Australia

Corresponding Author:
Ping Yu, PhD
Centre for Digital Transformation, School of Computing and Information Technology
Faculty of Engineering and Information Sciences
University of Wollongong
Northfields Avenue
Wollongong, 2522
Australia
Phone: 61 2 4221 5412
Fax: 61 2 4221 4170
Email: ping@uow.edu.au

Abstract

Background: Hypertension affects over 15% of the world’s population and is a significant global public health and socioeconomic challenge. Mobile health (mHealth) services have been increasingly introduced to support hypertensive patients to improve their self-management behaviors, such as adherence to pharmacotherapy and lifestyle modifications.

Objective: This study aims to explore patients’ perceptions of mHealth services and the mechanisms by which the services support them to self-manage their hypertension.

Methods: A semistructured, in-depth interview study was conducted with 22 outpatients of the General Hospital of Ningxia Medical University from March to May 2019. In 2015, the hospital introduced an mHealth service to support community-dwelling outpatients with self-management of hypertension. Content analysis was conducted by following a grounded theory approach for inductive thematic extraction. Constant comparison and categorization classified the first-level codes with similar meanings into higher-level themes.

Results: The patient-perceived mechanisms by which the mHealth service supported their self-management of hypertension were summarized as 6A: access, assessment, assistance, awareness, ability, and activation. With the portability of mobile phones and digitization of information, the mHealth service provided outpatients with easy access to assess their vital signs and self-management behaviors. The assessment results gave the patients real-time awareness of their health conditions and self-management performance, which activated their self-management behaviors. The mHealth service also gave outpatients access to assistance, which included health education and self-management reminders. Both types of assistance could also be activated by abnormal assessment results, that is, uncontrolled or deteriorating blood pressure values, discomfort symptoms, or not using the service for a long period. With its scalable use to handle any possible information and services, the mHealth service provided outpatients with educational materials to learn at their own pace. This led to an improvement in self-management awareness and ability, again activating their self-management behaviors. The patients would like to see further improvements in the service to provide more useful, personalized information and reliable services.

Conclusions: The mHealth service extended the traditional hypertension care model beyond the hospital and clinician's office. It provided outpatients with easy access to otherwise inaccessible hypertension management services. This led to process improvement for outpatients to access health assessment and health care assistance and improved their awareness and self-management ability, which activated their hypertension self-management behaviors. Future studies can apply the 6A framework to guide the design, implementation, and evaluation of mHealth services for outpatients to self-manage chronic conditions.

(J Med Internet Res 2021;23(6):e25522) doi:10.2196/25522

KEYWORDS

patient experience; mHealth; mobile phone; mobile app; intervention; self-management; high blood pressure; chronic disease management; qualitative research

Introduction

Background

Hypertension is one of the principal modifiable risk factors for cardiovascular diseases, especially for stroke and chronic renal diseases [1]. Without a clear pathophysiology or cure, it remains a major global burden of morbidity and mortality [1]. Alarmingly, rising living standards and their derived unhealthy lifestyles are accelerating the prevalence of hypertension annually. In 2015, an estimated 1.13 billion people worldwide had hypertension, leading to 19% of global deaths [2,3]. Hypertension also imposes a substantial economic burden on society and health care systems. The annual medical cost of hypertension is estimated at US $370 billion worldwide, representing approximately 10% of the global health care expenditure [4]. Thus, hypertension control is a priority for improving population health and containing chronic disease burden.

Hypertension treatment includes both pharmacological and nonpharmacological interventions [5]. These interventions are only effective if a patient engages in long-term, even lifelong, self-management behaviors, which include adherence to pharmacotherapy and lifestyle modifications. Patients need to take antihypertensive medication as often as prescribed and regularly measure blood pressure (BP) to monitor its efficacy [2,6]. In addition, they need to follow a healthy lifestyle, such as maintaining a healthy diet, performing physical activities, avoiding tobacco and unhealthy alcohol use, and managing mental stress [2,7]. However, for most patients, adherence to self-management behaviors is challenging. Less than 50% of patients adhere to hypertension self-management 1 year after initiating pharmacotherapy [8]. Nonadherent behaviors include failure to initiate pharmacotherapy and lifestyle modifications, take antihypertensive drugs at the prescribed frequency, adhere to long-term treatment, and monitor efficacy regularly [9,10]. These challenges can be addressed by assistive technology innovations, such as mobile health (mHealth) services.

mHealth Services

mHealth refers to the use of mobile devices to deliver medical or public health services [11]. As mHealth services break the constraints of time, space, cost, and capacity, they can provide patients with low-cost, affordable, ongoing support to manage their chronic conditions, such as hypertension, which requires long-term, lifestyle-related care plans and continuous monitoring in the home environment [12]. mHealth services have multiple modes, such as SMS text messages, mobile apps, and interactive voice responses [13]. With the proliferation of smartphones and the advancement of cellular networks, mHealth apps are increasingly available for patients to access health care services. In 2018, approximately 318,000 mHealth apps were available worldwide [14].

The effectiveness of mHealth services in assisting patients with self-management of hypertension has been widely studied. Empirical evidence suggests that mHealth services can better assist patients in controlling BP than the conventional face-to-face care model [15-20]. A review of 23 studies on the effectiveness of mHealth services found that 16 (69.6%) studies demonstrated a positive effect on medication adherence and healthy behavioral modification [21]. Lu et al [22] also found that mHealth apps can improve patient experience, especially in health information exchange, physician-patient communication, and short-term outcome improvement. Nevertheless, after assessing 186 hypertension-targeted apps in Apple and Google Play stores, Alessa et al [23] found that only 30 apps (16%) were likely to be effective in assisting hypertension self-management. The common feature of these 30 apps was that they all had 3 or more functionalities, including (but not limited to) self-monitoring, reminders and educational information or automatic feedback. There was little information about the theoretical basis of many apps, and there was no evidence about their effectiveness and usability. Song et al [24] also found that the ineffectiveness of the mHealth service might result from a lack of theoretical guidance for assisting behavioral modifications. Hallberg et al [25] interviewed 49 patients who used an interactive mobile phone–based system to self-manage hypertension. The patients appraised the system as a useful tool for self-reporting health conditions, measuring BP, retrieving self-reported data, and receiving motivational messages. These had led to improvements in lifestyle, health knowledge, and better engagement. However, as the participating patients were all active mobile app users, this study may be susceptible to positive bias. Conversely, Morrissey et al [26] questioned the sustainability of mHealth services despite positive patient engagement. To date, the mechanism for mHealth services to assist patients with self-management of hypertension is unclear.

Study Context and Aim

To pilot trial an mHealth service to improve patient hypertension management and population health, a large-scope, international, tripartite, collaborative program was conducted between the University of Wollongong, Australia; Zhejiang University, China; and the General Hospital of Ningxia Medical University, China, through a formal research partnership. The mHealth service entitled *BP Assistant* was developed by the Biomedical Informatics Laboratory at Zhejiang University to support patients in self-managing hypertension in Ningxia, a province in China with a high prevalence of hypertension because of high salt intake [19]. It was implemented in the General Hospital of Ningxia Medical University, the only tertiary hospital in the province, in November 2015. The hospital has more than 3000 beds and supports outpatients with hypertension self-management. In March 2019, 2079 patients were enrolled in the mHealth management program. This study explored patients' perceptions of the mHealth service and the service mechanisms to support them in self-managing hypertension.

Methods

Ethics Approval

This study was approved by the Human Research Ethics Committee of the General Hospital of Ningxia Medical University, China. The registration number is ID2018-325. Oral and written consent was obtained from all participants before the interviews, and their anonymity was preserved. All information and audio recordings of the participants were kept confidential.

Study Design, Setting, and Participants

This study was conducted using semistructured, in-depth interviews from March to May 2019 in the Department of Cardiology at the General Hospital of Ningxia Medical University. The study population included outpatients with hypertension who were registered to receive the mHealth service provided by this department. All participants had participated in a previous clinical trial (registration number ChiCTR1900026437) that evaluated the effectiveness of the service.

The mHealth Service

The mHealth service includes 2 technical components: a smartphone-based app, entitled BP Assistant, for patients to use to self-manage hypertension, and a web-based portal for health care providers to monitor and communicate with their patients. The app can be used in both iOS and Android systems. There are 6 key functional modules in the app: health education, health management plan, health checkup, health report, reminder service, and performance ranking (Textbox 1).

Textbox 1. Six key functional modules of the BP Assistant hypertension self-management app.

Health Education

- To provide educational information about hypertension and hypertension management

Health Management Plan

- To provide a to-do list, which requires users to record and upload personal health data, including blood pressure and heart rate, medication type and dose, weight, diet, exercise, and uncomfortable symptoms

Health Checkup

- To provide feedback to tell users whether their blood pressure is abnormal or normal

Health Report

- To provide visual daily and monthly reports based on a statistical analysis of input data

Reminder Service

- To allow users to set up and receive tailored reminders for health management plans

Performance Ranking

- To score and rank each user's performance according to their degree of app use in a ranking list

Health care providers can check all uploaded data from the web-based portal whenever needed. In particular, if an anomaly is detected, such as a sudden change in vital signs, long-term loss of contact, or patient-reported discomfort, the system will automatically send out a warning signal to alert the health care providers to follow up with the patient and provide treatment assistance or guidance.

Semistructured Interview Questions

The interview guide was developed by the first author in English and then translated into Chinese. Content validity was evaluated by a panel of 10 bilingual experts, including 4 cardiovascular medical specialists, 4 health informatics experts, 1 certified health manager, and 1 information systems expert, before translating back into English. Afterward, the interview guide was pilot tested for its face validity by 3 hypertensive patients who had used the app. Feedback from these patients was taken to further refine the interview guide to improve the understandability and relevance of each question (Multimedia Appendix 1).

Sampling and Recruitment

Participants were selected using purposive, snowball, and theoretical sampling until the data were saturated, that is, no new themes emerged [27-29]. We purposefully selected participants with varying sociodemographic characteristics to ensure diversity in age, gender, and type of mobile operating system being used. Theoretical sampling, which involves simultaneously collecting, coding, and analyzing data to determine whom to approach next to generate new insights, was used to enrich emerging categories and provide guidance for data collection to reduce selection bias.

Participants who (1) were aged 18 years or more, (2) had a diagnosis of essential hypertension (hypertension without identifiable causes) based on the Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure [5], and (3) had participated in the program to self-manage hypertension using the BP Assistant over 3 months were included in the study. Participants who (1) had secondary hypertension or (2) were unable to express their own perceptions because of mental disabilities or inability to speak were excluded from the study.

Potential participants were contacted by a registered nurse through a phone call or in a waiting room while they were waiting for a medical consultation. They were informed about the purpose and requirements of the study and then sought for verbal consent. An interview appointment was then made with those who provided verbal consent. At the beginning of the appointment, the interviewer described a detailed interview process and sought written consent. Only those who provided written informed consent were interviewed.

Data Collection

After providing written informed consent, the participant was guided into a reception room where the interview was conducted. After briefing the participant on the purpose, procedure, and content of the interview, the interview was started. Each interview lasted approximately 30 minutes and concluded when information saturation was reached. The audio recordings were transcribed verbatim in Microsoft Word documents. All names included in the manuscript were replaced by codes to preserve the anonymity and privacy of the participants. All information and audio recordings of the participants were kept confidential.

Data Analysis

The interview transcripts were systematically analyzed using an inductive thematic analysis method to identify different points of view and grouping them into various themes [30,31]. A total of 3 researchers engaged in content analysis. Each transcript was read carefully to obtain a general impression of the main experience of the participants by 2 researchers (TS and SQ). Content analysis followed the grounded theory approach [32,33]. The coding and classification were conducted in Microsoft Excel, benefiting from its affordance of easy visualization of a large number of codes in one screen, which was convenient for constant comparison, classification, and aggregation of codes, as suggested by Bree et al [34]. Each document was divided into sentences that only addressed one issue and were allocated to different cells in the same column. Each sentence was critically analyzed verbatim to identify the meaning unit by one researcher (TS), that is, an atomic text fragment containing relevant information to elucidate the research question. Given the different ways in which the participants expressed their experiences, each unit of experience was tagged with a label expressing its core meaning. This resulted in the abstraction of initial codes to describe these experiences.

The initial codes were compared, aggregated, and sorted into 16 higher conceptual–level concepts, known as subthemes. In this stage, how the patients were affected by the mHealth service was also established to understand the relationships. After that, concepts with similar meanings were further abstracted and grouped into an overarching 6 categories, called themes. After more than 3 months of constant comparison, aggregation, classification, and discussion, the tentative coding and data management were iteratively formulated and revised according to the consensus among the 4 researchers to avoid duplication and overlap (Table 1). The coding was then validated by another researcher (SQ), and the third researcher (PY) was referred to for arbitration when inconsistencies occurred.

Table 1. Examples of initial codes, subthemes, and themes from the interview.

Quotation	Initial code	Subtheme	Theme
“It is very convenient (when I am) outside (such as) going abroad. It can be taken and used everywhere.”	• Convenient to use outside • Convenient to use when going abroad • Can be taken and used everywhere	• Portability • Portability • Portability	• Access • Access • Access
“Just won't forget it. I used to forget about (taking medication on time). Now once the alarm rings, I will take the medications.”	• Do not forget • Take medicine on time when the alarm rings	• Reminder of self-management • Cue to action	• Assistance • Activation
“I used to think hypertension was not a problem. (But) the cases (provided in the app) awaked me. It is my own business to manage my daily life.”	• Not a problem • Cases • Own business	• Importance of self-management • Health education • Importance of self-management	• Awareness • Assistance • Awareness

Results

Demographics

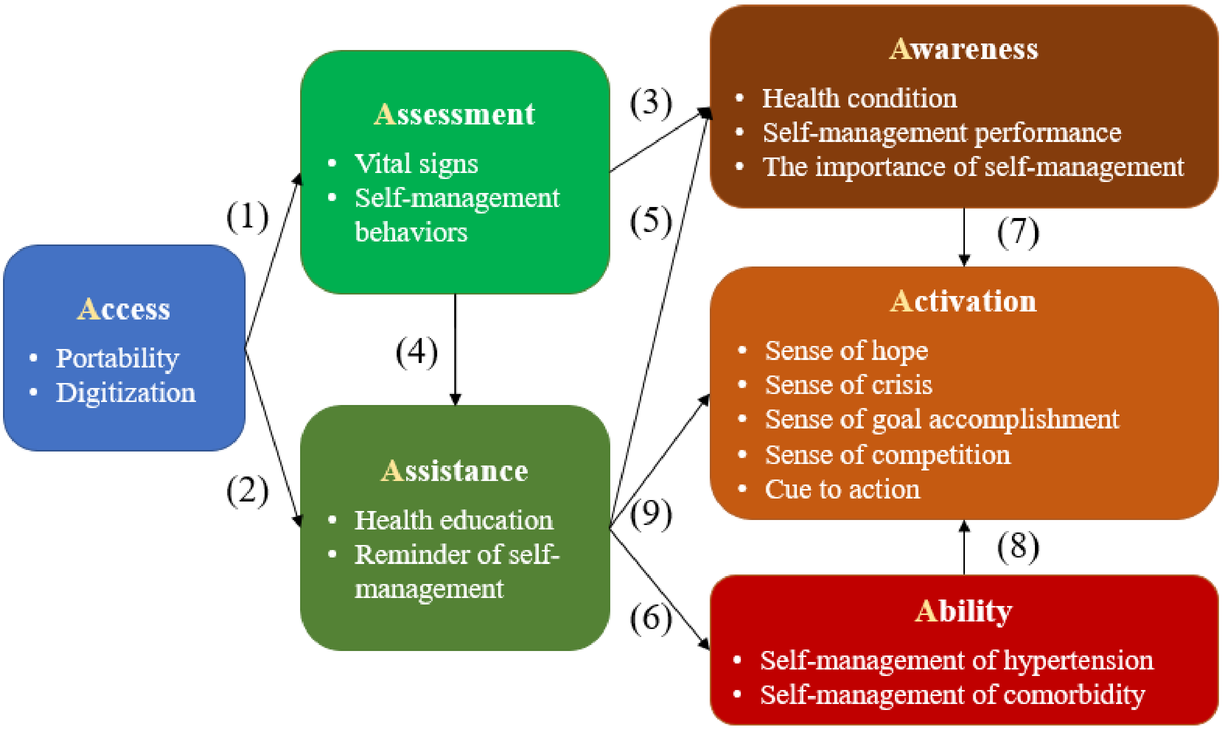
A total of 22 outpatients with hypertension (4/22 women, 18%) participated in this study. Their ages ranged from 33-73 years, with a median age of 47 years. They were employed (11/22, 50%), self-employed (7/22, 32%), retired (3/22, 14%), or unemployed (1/22, 5%). Of the total participants, 41% (9/22) had a bachelor's degree and above, 27% (6/22) had a college degree, 23% (5/22) held senior high school diplomas, and 9% (2/22) held junior high school diplomas. Out of 22 participants, 12 (55%) used iPhones, 8 (36%) used Android smartphones, and 2 (9%) used both.

A 6A Framework That Explains the Mechanism for the mHealth Service to Assist Patient in Self-Management of Hypertension

Overview

Patients' perceptions of the mHealth service on their hypertension self-management are summarized in 6 themes (6A): access, assessment, assistance, awareness, ability, and activation (Figure 1).

Figure 1. A 6A framework of using the mobile health service to assist patient in self-management of hypertension.



With the portability of mobile phones and digitization of information, the mHealth service provided hypertensive outpatients with (1) low-threshold *access* to health *assessment* and (2) health care *assistance*. (3) The *assessment* results, that is, vital signs and self-management behaviors, gave the patients real-time *awareness* of health conditions and self-management performance. (4) In particular, abnormal results, such as deterioration of vital signs or lack of self-management behavior records, could also be fed back to health care providers to provide the patients with *assistance*. Health care assistance included health education and reminders of self-management behaviors. (5) Health education assisted the patients in improving their *awareness* of the importance of self-management and (6) the *ability*, that is, knowledge and skills, in performing self-management behaviors. The reminders assisted patients in overcoming the barrier of forgetfulness. The improvement of (7) self-management awareness, (8) ability, and (9) reminders together *activated* their self-management behaviors.

In descending order of the number of times mentioned by the interview participants, the 16 subthemes were vital sign assessment, health education, ability to self-manage hypertension, digitization, reminder of self-management, awareness of the health condition, cue to action, portability, self-management behavior assessment, awareness of the importance of self-management, sense of goal accomplishment, sense of hope, sense of crisis, awareness of self-management performance, sense of competition, and ability to self-manage comorbidities (Multimedia Appendix 2).

Access

The patients appreciated that mHealth lowered their threshold to access health information and health care services compared with the conventional hypertension care model. This was because of the portability of mobile phones and digitization of information.

Portability

The use of the mHealth service for hypertension self-management was available through the smartphones that patients carried with them in daily life. The lightweight, portable smartphones made it convenient for them to access health assessments and health care assistance without time and space limitations. The patients could keep their health care providers informed of their health conditions through access to the app. Remote communication and interaction reduced the need for hospital visits, avoiding excessive registration and examination costs:

> *It is very convenient [when I am] outside [such as] going abroad. It can be taken and used everywhere.* [Patient 09]
>
> *It's easier for communication about hypertension management [with my doctor] because [if] there is always some tool to connect between us to exchange information, [I] do not have to run to the hospital. It has always been annoying to go to the hospital to seek medical treatment. We have to wait in a crowded space for a long time, and the queue is often quite long. Health problems that could be solved in a few minutes always took a whole morning.* [Patient 18]

Digitization

The mHealth service provided patients with a convenient digital platform to access health information and health care services. The data were in digital form and stored in the app, which was easier to retrieve and less likely to be lost than the traditional method of recording data in a notebook. It was also quicker and easier to review the records without flicking through paper-based records, allowing both patients and their health care providers to capture useful information in time:

> *In every record, the specific time was recorded; the specific data was also recorded...unlike writing it down in a notebook, it won't be lost.* [Patient 06]
>
> *It's so convenient that I can track my records easily,...,unlike before, if I needed to identify a certain record, I had to go through all written records one by one.* [Patient 14]
>
> *(It is) more convenient. I used to record my BP in a notebook, every time the doctor had to flip through it and read the records one by one. You see, it is impossible for her to read each record and identify my highest systolic BP. As she has to see so many patients in a day, in fact, she cannot get anything (meaningful)...(But) now she only needs to glance at the curve, the overall trend, the peaks and valleys directly on the mobile phone. This is faster and more intuitive.* [Patient 02]

Assessment

The patients praised the mHealth service because it provided them with the opportunity to assess their vital signs and self-management behavior.

Vital Signs

The mHealth service assessed the vital signs of patients, that is, BP and heart rate, measured and entered by themselves on the app through automatic analysis, with the results presented back to both the patients and their health care providers. The patients believed that these results provided evidence that enabled their health care providers to understand their condition, make an appropriate diagnosis, and provide feedback, such as adjusting the treatment program, instead of relying on a single measurement of BP in the clinic:

> *[The app] helped me to tweak my medicine well,...My medication had been changed four times before because my BP reading was always up and down. I did not have any records at that time...With the app, the doctor prescribed the medicine in reference to all the records of my BP. Once the medication was right, the BP was gradually under control.* [Patient 11]

One patient specifically pointed out that the mHealth service helped to solve his *white-coat* effect in the clinic:

> *...my BP readings are always higher when I sit with the doctor but lower at home. So, the data I upload [to the app] is more accurate than that taken in the hospital.* [Patient 09]

Self-Management Behavior

The mHealth service automatically recorded the patients' self-management behaviors, including BP, heart rate, and weight measurement; medicine intake; step count; and food intake, to assess their self-management performance. The patients thought that the assessment results provided their health care providers with evidence to determine the reason for the suboptimal efficacy, that is, whether the medicine was inappropriate or the patients did not adhere to the self-management:

> *In the past, if the BP was not well controlled, the doctors would ask me if I take medicine on time or if there was too much salt intake or I did not do exercise. With the record, she can immediately see what might cause poor BP control.* [Patient 17]

In particular, abnormal assessment results, that is, uncontrolled or deteriorating BP values, discomfort symptoms, or not using the system for a long period could also be detected by the mHealth system. Once an abnormal result was detected, the mHealth system would send an alarm signal to notify the health care provider to contact the patient or their families to provide timely feedback:

> *...[The health manager] called me in time and asked me what happened recently; why wasn't my BP controlled but suddenly increased? Did I eat high cholesterol food etc...[it] is a wake-up call to us.* [Patient 07]

The negative experiences reported by the patients included overwhelming assessment items, inaccurate assessment of health conditions, and inability to see how lifestyle changes impact BP levels. Some patients complained that the app required too many data items to be entered on a daily basis; they doubted about the actual effect on disease management but saw it as a waste of their time:

> *Some functions are useless. For example, I take the same drugs every day. Why do I need to upload them every day?* [Patient 12]
>
> *I only use [hypertension-]related functions. Also, it's hard to assess, i.e., the diet. For example, if I record a bowl of rice, you still don't know exactly how much rice I eat.* [Patient 07]

Some patients considered that some wordings in the app, such as BP and discomfort symptoms, were not entirely accurate:

> *Your assessment of discomfort symptoms is inaccurate and illogical. The symptoms listed in the app, such as chest pain and mixed-up words, are typical symptoms of cardio or cerebral infarction. How can I still stay at home and wait for feedback once they occur? I must have to directly go to the hospital or call the emergency. If you say sweating, a little dizzy, these can be called discomfort symptoms.* [Patient 02]
>
> *The assessment of BP is unscientific without considering the pressure difference. I remember one time, I felt particularly uncomfortable, and my BP was 100/80 [mmHg]. The app showed unexpectedly that my BP was normal. This would mislead the patient.* [Patient 12]

One patient strongly felt a lack of logical connection between the assessment results, for example, BP values and other data items. This impeded a patient's recognition of the importance of certain app functions for hypertension management, thus reducing their acceptance and use of these functions:

> *The assessment indicators [in the app] are all isolated. We can't see any connection between what we do and the corresponding outcomes.* [Patient 09]

Assistance

The patients reported 2 types of assistance provided by the mHealth service: health education and reminders.

Health Education

The patients acknowledged that their physicians had explained, more or less, the reason and the way to self-manage hypertension; however, they either did not fully understand or ignored these instructions during the medical consultation period. The mHealth service assisted them with this challenge because it provided a large volume of educational materials that were not otherwise accessible, which allowed them to learn at their own pace so as to absorb the information and to truly understand the meaning of the medical orders:

> *Yes, [the doctor] told me before. But, he or she may not go into the level of detail due to time limit. (I) didn't take it to heart either [because I' only paid attention to what medications were (at the time of medical consultation).* [Patient 18]
>
> *...the doctor had told me everything [about self-management of hypertension] when I was sitting with her. But I had little knowledge and did not really absorb the information. When encountering the same information again [in the app], I came to understand why and how to [do it].* [Patient 03]

The patients also felt that the content provided by the mHealth service was more trustworthy than those obtained from other sources because it was disease-focused and recommended by health care professionals in a prestigious tertiary hospital:

> *...although it is much easier to get the [hypertension-related] information from other sources, such as TV, mobile phone, newspapers, your app is more trustworthy because it is targeted, specifically for hypertension and it's sophisticated. Also, [it is] professional because [the content is] checked by doctors, unlike information [from other sources which was] just copied and pasted [from somewhere else]. [They] cannot be trusted.* [Patient 02]

Reminders

The mHealth service reminded them to perform necessary self-management actions, such as taking medications, which assisted them in overcoming the barrier of forgetfulness:

> *Just won't forget it. I used to forget about [taking medication on time]. Now once the alarm rings, I will take the medications.* [Patient 09]

The negative experiences mentioned included that the feedback service was lagged behind and untimely because they had experienced symptoms and discomfort without receiving any feedback:

> *...I felt chest pain at one time, and I reported it in the app. But there was no response.* [Patient 22]

There were also complaints about the technical instability of the app, preventing the patient from getting timely assistance:

> *Sometimes I can't log into my account. You know, users will lose confidence in your product in this case.* [Patient 15]
>
> *Occasionally, it may fall offline.* [Patient 13]

Awareness

The patients felt that their awareness was improved by using the mHealth service, which included awareness of their health conditions and self-management performance and the importance of self-management.

Health Conditions and Self-Management Performance

The visual and colorful graphic assessment results charts for vital signs and self-management behaviors improved patients' awareness of their health conditions and self-management performance, respectively:

> *...then translate [the assessment results] to a graph. I was more aware [about my health condition] as soon as I looked at the chart. My BP, for example, showed a downward trend but had not met the target.* [Patient 02]
>
> *It has several indicators, in different colors, it's very vivid, clear at a glance.* [Patient 05]

Importance of Self-Management

Despite being diagnosed with hypertension, some patients never took it seriously because they did not believe in the consequences of the disease or the need to control it. The patients admitted that the educational materials assisted them with a better understanding of hypertension-related risk factors and corresponding consequences and built their belief in the responsibility of self-management. This made them change their attitudes, being aware of the importance of self-management, and they thus had the intention to engage in self-management behaviors:

> *When diagnosed with hypertension after a medical examination, I thought the doctor just scared me because I didn't feel anything wrong with my health. I ate and drank and smoked whatever and whenever I wanted, and even didn't take any medication because I was afraid of side-effect. After reading the materials in the app, I knew I was wrong. It was time to manage hypertension.* [Patient 19]
>
> *I have never managed [hypertension] before [using the app because] I thought many of my family and friends also had [it], but they were still alive and well. Now I know that once there are symptoms, it is late because our organs have been injured, and the injury was mostly irreversible.* [Patient 04]
>
> *I used to think hypertension was not a problem. If [I felt] dizzy, [then I] go to see a doctor; otherwise, just let it be. [But] the cases [provided in the app] awaked me. It is my own business to manage my daily life.* [Patient 18]

Ability

Patients reported that using the mHealth service improved their ability to self-manage hypertension, including their knowledge and skills in hypertension self-management and inspiration for self-management of comorbidities.

Knowledge and Skills in Self-Management of Hypertension

The health education provided by the mHealth service increased patients' general knowledge of hypertension self-management and improved their BP control skills and strategies for practicing healthy lifestyles:

> *The health tips are quite useful, [I] got to know what should [I] pay attention to in the daily life, and how to achieve self-regulation; for example, I often remind myself not to be angry, to maintain a healthy diet, and to do physical activities.* [Patient 15]
>
> *...I went through the educational material little by little, such as how to measure BP correctly at home, what matters when taking antihypertensive drugs, etc. Now, I know how to self-manage [hypertension] well.* [Patient 01]

In particular, one patient mentioned that the mHealth service improved his ability to deal with discomfort symptoms. As he learned about the signs and patterns of the manifested symptoms and the level of severity, he was confident and capable of deciding whether to wait and see or to seek medical assistance immediately:

> *...The educational materials are quite useful...[For example, previously] I always rushed to the emergency department when I felt dizzy or heart palpitations because I had no idea what happened. But now what I usually do is taking out a gauge to measure my BP and heart rate, if they are abnormal, I will observe it for a while. Usually, they would return to normal without doing anything. Otherwise, I would see the doctor. In addition, I used to be very anxious when this case happened, which made my BP higher; but now, I always remind myself to be calm, making a rational decision after observation.* [Patient 05]

Inspiration for Self-Management of Comorbidities

The mHealth service also improved patients' ability to self-manage comorbidities because there were similar requirements for managing chronic diseases:

> *Although this [app] is hypertension-focused, it is virtually enlightening to my daily self-management of coronary heart disease and diabetes, such as eating, exercising, medicine adherence, because their management principles are similar and interlinked.* [Patient 09]

However, some patients believed that the information provided by the app was not sufficient to improve their ability because the content was monotonous and difficult to absorb and the update was not frequent enough:

> *The content of health education is monotonous, and lack regular updates.* [Patient 07]
>
> *The health education material was not adequate, thus less useful. The information was out of date.* [Patient 14]

Activation

The patients reported that using the mHealth service activated their self-management behaviors, which included their sense of hope, sense of crisis, sense of goal accomplishment and sense of competition, and cues to action.

Sense of Hope and Crisis

The assessment results of vital signs gave the patients real-time awareness of their health conditions, thereby activating their sense of hope or crisis. Awareness of improved health conditions provided a sense of hope for BP control, which encouraged them to maintain self-management behaviors:

> *It brought up a sense of achievement by seeing the curve of my BP recording being flattened. This gives me the motivation to keep going. You see, [my systolic BP] has dropped from 164 [mmHg] to 142 [mmHg]. It will be normal once [it] drops to 120 [mmHg].* [Patient 05]

Conversely, awareness of deteriorating health conditions provided a sense of crisis, which activated them to improve self-management strategies:

> *A curve shown in the app can indicate whether my heart rate is fast, slow or normal today, like my BP. If abnormal, I was set on alert to reflect upon my own behavior. Should I get more exercise? Am I eating too much salt? Did I forget to take my medicine?* [Patient 14]

> *...if you don't record [the BP], there is no [data] for comparison...But if you have recorded and compared yesterday's, today's and tomorrow's records, a sense of crisis will push you to reflect. [For example], my BP is rising. What might have caused it? Oh, probably because I drank alcohol. The record showed my BP was normal before. So I shouldn't drink alcohol in a couple of days; instead, I should pay attention to a healthy diet, regular work and rest.* [Patient 15]

Sense of Goal Accomplishment and Competition

The assessment results of self-management behaviors made the patients aware of self-management performance of both themselves and other users, thereby activating their sense of goal accomplishment and competition. Persistent recording of patients' self-management efforts had brought them a sense of goal accomplishment:

> *...it would be nice that my effort could be faithfully recorded by your app. I did it without any recording in the past. But now, look at these records. I have the strength [to keep going].* [Patient 15]

> *What I did and not have been recorded [in the app]. For example, walking, I was looking through [the records] now and then. [If] I walked 9,000 steps yesterday, I knew I reached the standard and felt at ease. If I did not reach 9000 steps today, I always worried and thought that I should go out and walk more.* [Patient 21]

In particular, 2 patients mentioned that the performance-ranking function enabled activation. When a patient knew that other patients did better than they had done, their sense of competition was switched on, which activated self-management behaviors:

> *Seeing that someone was ahead of me, I made a firm decision that I had to catch up with him or her.* [Patient 16]

Cues to Action

The reminder activated patients' self-management behaviors by providing them with prompts to perform a specific action, for example, taking medications:

> *...[the app] always reminds me to take medicine on time.* [Patient 18]

> *[Previously,] the doctor couldn't track my health conditions, and I wasn't completely sure [about it]...so I was anxious. But now, I was immediately told that I should go to the hospital and take a look at it, which made me felt safe. I felt that I was put on the radar screen. They had been paying close attention to me with timely feedback, so I should also play an active role to manage myself well.* [Patient 03]

Despite the perceived benefits of self-management activation, certain patients suggested further improvement in the incentivizing mechanism for using mHealth services:

> *Actually, [if] you want to encourage them, you have to send a message to list their achievement, such as how many tasks have been completed, how well they controlled [their BP], and how long they have persisted in using the app, etc.* [Patient 07]

> *...what can I get? For example, I use another app, and if I walk up to 5,000 steps, it will pop up a message saying, "Congratulations! You have reached 5,000 steps today. Please keep it up". I can't see recognition of what I achieved using this app.* [Patient 09]

> *I don't know the criteria for scoring and ranking the performance, so I'm not motivated because I don't know what to do next.* [Patient 15]

Discussion

Principal Findings

Compared with the traditional hypertension care model, the mechanism for the mHealth service to support patient self-management of hypertension can be summarized in a 6A framework: access, assessment, assistance, awareness, activation, and ability (Figure 2).

Our findings can be explained by Donabedian health care quality model [35]. The mHealth service changes the structure, that is, the delivery mode and process of the traditional hypertension care model. It provides patients with easy access to health assessment and health care assistance, which assists with improved self-management awareness and ability. These again activate self-management behaviors (Figure 3).

Figure 2. The mechanism for the mobile health service to support patients to self-manage hypertension.

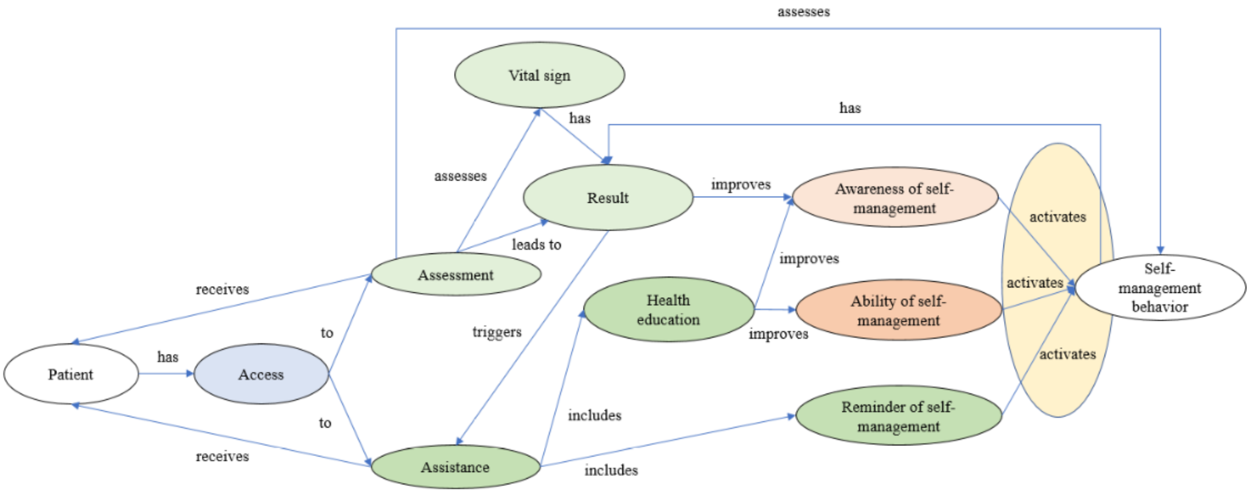
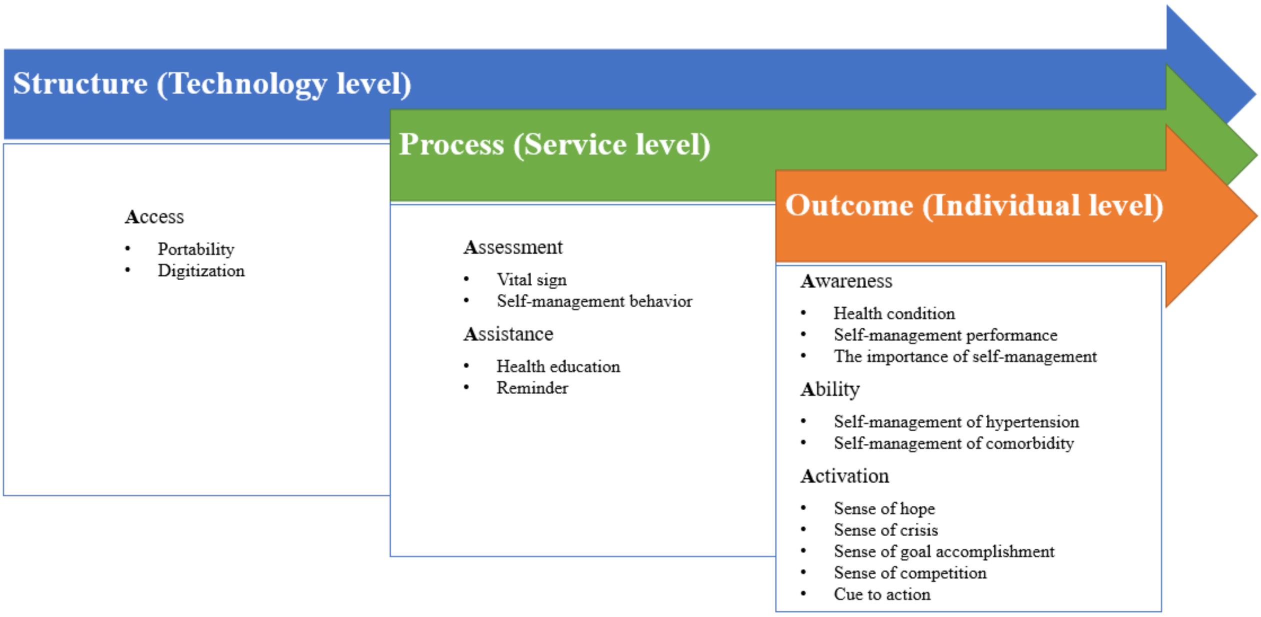


Figure 3. Interactions among the structure (access), process (assessment and assistance), and outcome (awareness, ability, and activation) in the mobile health service for hypertension self-management.



The mHealth service enables outpatients to access health care services that were previously only available to inpatients. These included the assessment results of vital signs and self-management behaviors and assistance from the mHealth system and health care providers. The increased access is attributed not only to the functions of the app but also to the changes in access rights facilitated by the mHealth delivery mode. Computer automation reduces labor costs to facilitate structural changes in the scope of health care service delivery beyond the boundaries of hospitals. This change in the health care service structure appears to lead to improved clinical processes [24]. Our patients perceive app use, or smartphone use, as an effective and easy way of gaining access to health care services, facilitating their acceptance and use of mHealth apps, as found by Anderson et al [36].

Smartphone portability encourages patients to interact with the app and extensively use app functions without geographical boundaries and time restrictions [37,38]. The patients also appreciate that the mHealth service lowers the threshold for them to access health care services. They do not need to make a long journey and wait in a long queue to see their health care providers, saving them time and cost.

The mHealth service also improves data access through timely data transfer and easier data retrieval [39-41]. Vital sign data are presented in easy visual graphics, allowing patients to read the information on a mobile phone at a glance, instead of reading it on paper or a computer. As suggested by Hallberg et al [25], if the graph can be viewed directly on a mobile phone without connecting to a computer, the mHealth service can be more useful because it addresses the barrier to using the system.

A total of 2 types of processes are changed—assessment and assistance. Patients have access to assessing their vital signs and self-management behaviors and the assistance of health education and reminders at home.

The assessment results improve patients' self-awareness of their health conditions, that is, *What is my current situation*? It also provides evidence to health care providers for accurate and timely diagnosis and prescription, as found by Myers et al [42]. However, Den Hond et al [43] stressed that self-measurement of BP cannot replace ambulatory monitoring for clinical diagnosis. Interestingly, the *white-coat effect*, that is, rising BP when seeing a doctor, can distort the real health condition of the patients and lead to misdiagnosis and treatment in face-to-face consultations [44]. Regularly measuring BP at home and recording the results on the app can help solve this problem. This is in line with the findings of Barsky et al [45]. It is worth noting that some patients reported low accuracy of BP assessment results. If there is a false assessment result, it will inevitably affect the judgment of the severity of the disease and even lead to wrong actions. This indicates that future product development should involve medical experts to ensure assessment accuracy [46].

As suggested by Norman Cousins, "Each patient carries his own doctor inside him" [47]; self-management is essential for all patients with chronic conditions, such as hypertension. The first step was to establish awareness and ability. Lack of awareness of the importance of self-management is a major factor causing poor BP control [48]. Bokhour et al [49] also suggested that hypertensive patients' awareness of factors that affect BP in their daily life may influence their ability to perform self-management behaviors. Our findings suggest that health education has improved patients' awareness of health self-management by addressing their *why* questions, that is, *Why is it important to self-manage hypertension?* Despite best efforts, limited medical resources and time make it difficult to educate patients and meet their individual needs in traditional face-to-face interventions [50]. The advantage of the mHealth service comes from its high scalability for providing services, for example, disseminating a large amount of educational material to a wide range of patients with hypertension for effective health education for chronic disease management [51-53]. It also allows these patients to absorb the information in their own time and at their own pace until they are fully aware of the importance and have the ability to perform self-management. They can also continuously improve their self-management abilities through constant learning. This is in accordance with the basic requirements for successful pharmacotherapy and lifestyle management for hypertension. However, some patients were disappointed with the publication frequency, accuracy, and legibility of the educational content. This suggests that a successful mHealth service requires accurate resources to be published regularly to sustain patients' trust with the mHealth service, which was in accordance with the study by Zarea et al [54] to evaluate health information websites. The negative feedback from the patients also provided insights into what and why the mHealth services did not work, that is, violating the 6A mechanism.

Awareness of one's health conditions and self-management performance can activate a person to perform self-management behaviors [55]. This is likely to be achieved by stimulating their sense of hope, sense of crisis, sense of goal accomplishment, and sense of competition, which are the internal driving forces for behaviors. Deteriorating BP changes can enable patients to take responsibility for managing their own health, activating the urge to perform self-management behaviors. This can be explained by the protection motivation theory, which states that protective motivation is affected by the perceived severity of the disease [56]. A similar phenomenon was found in patient self-management of chronic kidney disease, chronic fatigue syndrome, and unhealthy alcohol use [57-59].

Forgetting to perform self-management is a common problem facing patients with hypertension in daily life. The automatic reminder provides a prompt in solving this problem. It serves as an external driving force to complement the internal driving forces mentioned above. This is in line with the health belief model, suggesting that a *cue* is essential to guide people to perform health-promoting behaviors, especially for the older people [13,24,46,60].

Our results also corroborate and complement the existing behavior change theories, such as the COM-B (Capability, Opportunity, Motivation, Behavior) model proposed by Michie [61]. The Capability, Opportunity, Motivation, Behavior model describes the interactions between 3 components—capability (C), opportunity (O), and motivation (M)—which produce behaviors (B) that, in turn, affect these components. *Capability* is defined as the individual's mental and physical capacity to engage in related activities. It is similar to *ability* in our 6A framework, that is, the patient's knowledge and skills to perform self-management. *Opportunity* is defined as all external factors that make the behavior possible or prompt it. It corresponds to access, assessment, and assistance provided by the mHealth service in the 6A framework. *Motivation* is defined as all brain processes that can stimulate and guide behavior, similar to *activation* in the 6A framework, which includes habitual processes (cue to action) and emotional responses (sense of hope or sense of crisis or sense of accomplishment or sense of competition). One step further, the 6A framework highlights the concept of *awareness*. This is supported by another behavioral change theory, the Theory of Planned Behavior. According to the Theory of Planned Behavior, the premise of a behavior change is the intention to change [62]. In contrast, internal capabilities and external opportunities are not adequate conditions for behavioral change. In the patients' stories, it was repeatedly mentioned that the driving force for them to engage in mHealth service is *awareness* of the need to change and the mechanisms to make change. Therefore, improving self-*awareness* is one of the key mechanisms for mHealth services to function.

Limitations

Our findings were drawn from interviews of 22 patients to understand their subjective perceptions of the mHealth service. Therefore, the findings were influenced by their own experience, despite the fact that studying patients' perceptions is a rational approach for understanding information system performance

[39]. Confinement of the study to a single hospital may limit the generalizability of the results to other settings. There is a limitation in translating the results of a mobile app intervention for hypertension management to other mHealth services media, such as text messaging and interactive voice response [13,59], or for other chronic diseases, such as diabetes. Different apps can have different functions and delivery formats. This suggests the need for further studies on patients' needs and preferences to understand which functions and mode of delivery are most helpful to them. Moreover, the analysis can be difficult to be completely free from a priori views and knowledge of the hypertension management system of the authors, an inevitable limitation of qualitative content analysis [33]. However, each sentence in the transcripts was read verbatim to understand the meaning of each participant as neutral and accurately as possible. The verification of the analysis by 2 more researchers also made up for this limitation as much as possible.

Conclusions

The mHealth service extended the structure of the traditional hypertension care model beyond the hospital and clinician's office. With its portability and digitization, the mHealth service provided patients with low-threshold access to communicating with their health care providers and receiving health care services to support their self-management of hypertension at home. Such structural changes in health service delivery have brought process improvement to assist patients' access to effective health assessment and health care assistance anytime, anywhere. The improvement in awareness and self-management ability and reminders brought about by such structural and process changes activated their hypertension self-management behaviors. They would like to see the mHealth service to provide more useful functions and easy-to-use services. Therefore, the comprehensive 6A framework extracted from the in-depth qualitative research theorized the mechanism for the mHealth service to improve patient self-management of hypertension. The mechanism can be further applied to guide the design, implementation, and evaluation of mHealth services for outpatients to self-manage chronic conditions.

Acknowledgments

The authors would like to thank all the patients for their participation in the interview and Diane Walton for proofreading the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[PDF File (Adobe PDF File), 61 KB - jmir_v23i6e25522_app1.pdf]

Multimedia Appendix 2
The number of times each subtheme was mentioned by the participants in the interviews.
[PNG File , 40 KB - jmir_v23i6e25522_app2.png]

References

1. Olsen MH, Angell SY, Asma S, Boutouyrie P, Burger D, Chirinos JA, et al. A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the Lancet Commission on hypertension. Lancet 2016 Dec 26;388(10060):2665-2712. [doi: 10.1016/S0140-6736(16)31134-5] [Medline: 27671667]
2. Hypertension. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/hypertension [accessed 2021-06-04]
3. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, et al. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990-2015. J Am Med Assoc 2017 Dec 10;317(2):165-182. [doi: 10.1001/jama.2016.19043] [Medline: 28097354]
4. Gaziano TA, Bitton A, Anand S, Weinstein MC, International SO. The global cost of nonoptimal blood pressure. J Hypertens 2009 Jul;27(7):1472-1477. [doi: 10.1097/HJH.0b013e32832a9ba3] [Medline: 19474763]
5. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. National Heart, Lung, and Blood Institute, National High Blood Pressure Education Program Coordinating Committee. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. Hypertension 2003 Dec;42(6):1206-1252 [FREE Full text] [doi: 10.1161/01.HYP.0000107251.49515.c2] [Medline: 14656957]
6. Magid DJ, Farmer SA. Hypertension self-management: a home run for patients and payers. Circ Cardiovasc Qual Outcomes 2014 Mar;7(2):205-206. [doi: 10.1161/CIRCOUTCOMES.114.000885] [Medline: 24619316]

7. Jafari F, Shahriari M, Sabouhi F, Khosravi-Farsani A, Eghbali-Babadi M. The effect of a lifestyle management educational program on blood pressure, heart rate, and body mass index in patients with hypertension. ARYA Atheroscler 2020 Jan;16(1):7-15 [FREE Full text] [doi: 10.22122/arya.v16i1.1502] [Medline: 32499826]
8. Hill MN, Miller NH, Degeest S, American Society of Hypertension Writing Group, Materson BJ, Black HR, et al. Adherence and persistence with taking medication to control high blood pressure. J Am Soc Hypertens 2011;5(1):56-63. [doi: 10.1016/j.jash.2011.01.001] [Medline: 21320699]
9. Burnier M, Egan BM. Adherence in hypertension: a review of prevalence, risk factors, impact, and management. Circ Res 2019 Mar 29;124(7):1124-1140. [doi: 10.1161/circresaha.118.313220]
10. Redon J, Brunner HR, Ferri C, Hilgers KF, Kolloch R, van Montfrans G. Practical solutions to the challenges of uncontrolled hypertension: a white paper. J Hypertens Suppl 2008 Dec;26(4):1-14. [doi: 10.1097/01.hjh.0000343507.74401.45] [Medline: 19096292]
11. World Health Organization. mHealth: New Horizons for Health through Mobile Technologies: Second Global Survey on eHealth (Global Observatory for eHealth series - Volume 3). Geneva: World Health Organization; 2011:1-111.
12. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
13. Song T, Yu P. mHealth. In: Gu D, Dupre ME, editors. Encyclopedia of Gerontology and Population Aging. Switzerland: Springer International Publishing; 2019:1-7. [doi: 10.1007/978-3-319-69892-2_452-2]
14. The rise of mHealth apps: a market snapshot. Liquid State. 2018. URL: https://liquid-state.com/mhealth-apps-market-snapshot/ [accessed 2021-06-04]
15. Mann DM, Kudesia V, Reddy S, Weng M, Imler D, Quintiliani L. Development of DASH Mobile: a mHealth lifestyle change intervention for the management of hypertension. Stud Health Technol Inform 2013;192:973. [Medline: 23920747]
16. Kang H, Park H. A mobile app for hypertension management based on clinical practice guidelines: development and deployment. JMIR Mhealth Uhealth 2016 Feb 02;4(1):e12 [FREE Full text] [doi: 10.2196/mhealth.4966] [Medline: 26839283]
17. Grant S, Hodgkinson J, Schwartz C, Bradburn P, Franssen M, Hobbs FR, et al. Using mHealth for the management of hypertension in UK primary care: an embedded qualitative study of the TASMINH4 randomised controlled trial. Br J Gen Pract 2019 Sep;69(686):612-620 [FREE Full text] [doi: 10.3399/bjgp19X704585] [Medline: 31262847]
18. Moon EW, Tan NC, Allen JC, Jafar TH. The use of wireless, smartphone app-assisted home blood pressure monitoring among hypertensive patients in Singapore: pilot randomized controlled trial. JMIR Mhealth Uhealth 2019 May 28;7(5):e13153 [FREE Full text] [doi: 10.2196/13153] [Medline: 30905872]
19. Lu X, Yang H, Xia X, Lu X, Lin J, Liu F, et al. Interactive mobile health intervention and blood pressure management in adults. Hypertension 2019 Sep;74(3):697-704. [doi: 10.1161/HYPERTENSIONAHA.119.13273] [Medline: 31327259]
20. Santo K, Redfern J. The potential of mHealth applications in improving resistant hypertension self-assessment, treatment and control. Curr Hypertens Rep 2019 Oct 09;21(10):81. [doi: 10.1007/s11906-019-0986-z] [Medline: 31598792]
21. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
22. Lu C, Hu Y, Xie J, Fu Q, Leigh I, Governor S, et al. The use of mobile health applications to improve patient experience: cross-sectional study in Chinese public hospitals. JMIR Mhealth Uhealth 2018 May 23;6(5):e126. [doi: 10.2196/mhealth.9145] [Medline: 29792290]
23. Alessa T, Hawley MS, Hock ES, de Witte L. Smartphone apps to support self-management of hypertension: review and content analysis. JMIR Mhealth Uhealth 2019 May 28;7(5):e13645 [FREE Full text] [doi: 10.2196/13645] [Medline: 31140434]
24. Song T, Qian S, Cui T, Yu P. The use of theory in mobile health interventions for patient self-management of chronic diseases. Stud Health Technol Inform 2019 Aug 21;264:1982-1983. [doi: 10.3233/SHTI190745] [Medline: 31438439]
25. Hallberg I, Ranerup A, Kjellgren K. Supporting the self-management of hypertension: patients' experiences of using a mobile phone-based system. J Hum Hypertens 2016 Feb;30(2):141-146 [FREE Full text] [doi: 10.1038/jhh.2015.37] [Medline: 25903164]
26. Morrissey EC, Casey M, Glynn LG, Walsh JC, Molloy GJ. Smartphone apps for improving medication adherence in hypertension: patients' perspectives. Patient Prefer Adherence 2018;12:813-822 [FREE Full text] [doi: 10.2147/PPA.S145647] [Medline: 29785096]
27. Draucker CB, Martsolf DS, Ross R, Rusk TB. Theoretical sampling and category development in grounded theory. Qual Health Res 2007 Oct;17(8):1137-1148. [doi: 10.1177/1049732307308450] [Medline: 17928484]
28. Noy C. Sampling knowledge: the hermeneutics of snowball sampling in qualitative research. Int J Soc Res Methodol 2008 Oct;11(4):327-344. [doi: 10.1080/13645570701401305]
29. Gentles S, Charles C, Ploeg J, McKibbon KA. Sampling in qualitative research: insights from an overview of the methods literature. Qual Rep 2015 Nov 7;11(4):327-344. [doi: 10.46743/2160-3715/2015.2373]

30. Martin PY, Turner BA. Grounded theory and organizational research. J Appl Behav Sci 2016 Jul 26;22(2):141-157. [doi: 10.1177/002188638602200207]
31. Kolb SM. Grounded theory and the constant comparative method: valid research strategies for educators. J Emerg Trends Edu Res Policy Stud 2012;3(1):83-86 [FREE Full text] [doi: 10.10520/EJC135409]
32. Strauss A, Corbin J. Grounded theory methodology. In: Denzin NK, Lincoln YS, editors. Handbook of Qualitative Research. Thousand Oaks, California, United States: Sage Publications Inc; 1994:273-285.
33. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
34. Bree R, Gallagher G. Using Microsoft Excel to code and thematically analyse qualitative data: a simple, cost-effective approach. All Ireland J High Edu 2016;8(2):A [FREE Full text]
35. Donabedian A. Evaluating the quality of medical care. 1966. Milbank Q 2005;83(4):691-729 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00397.x] [Medline: 16279964]
36. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
37. Toro-Ramos T, Kim Y, Wood M, Rajda J, Niejadlik K, Honcz J, et al. Efficacy of a mobile hypertension prevention delivery platform with human coaching. J Hum Hypertens 2017 Dec;31(12):795-800 [FREE Full text] [doi: 10.1038/jhh.2017.69] [Medline: 28972573]
38. Agarwal S, Lau CT. Remote health monitoring using mobile phones and web services. Telemed J E Health 2010 Jun;16(5):603-607. [doi: 10.1089/tmj.2009.0165] [Medline: 20575728]
39. Kim DJ, Hwang Y. A study of mobile internet user's service quality perceptions from a user's utilitarian and hedonic value tendency perspectives. Inf Syst Front 2010 Sep 7;14(2):409-421. [doi: 10.1007/s10796-010-9267-8]
40. Zhang Y, Yu P, Shen J. The benefits of introducing electronic health records in residential aged care facilities: a multiple case study. Int J Med Inform 2012 Oct;81(10):690-704. [doi: 10.1016/j.ijmedinf.2012.05.013] [Medline: 22749424]
41. Ammenwerth E, Buchauer A, Bludau B, Haux R. Mobile information and communication tools in the hospital. Int J Med Inform 2000 Jan;57(1):21-40. [doi: 10.1016/s1386-5056(99)00056-8] [Medline: 10708253]
42. Myers MG. Recent advances in automated blood pressure measurement. Curr Hypertens Rep 2008 Oct;10(5):355-358. [doi: 10.1007/s11906-008-0067-1] [Medline: 18775111]
43. Hond ED, Celis H, Fagard R, Keary L, Leeman M, O'Brien E, THOP investigators. Self-measured versus ambulatory blood pressure in the diagnosis of hypertension. J Hypertens 2003 Apr;21(4):717-722. [doi: 10.1097/00004872-200304000-00014] [Medline: 12658017]
44. Spruill TM, Pickering TG, Schwartz JE, Mostofsky E, Ogedegbe G, Clemow L, et al. The impact of perceived hypertension status on anxiety and the white coat effect. Ann Behav Med 2007 Aug;34(1):1-9. [doi: 10.1007/BF02879915] [Medline: 17688391]
45. Barsky J, Hunter R, McAllister C, Yeates K, Campbell N, Liu P, et al. Analysis of the implementation, user perspectives, and feedback from a mobile health intervention for individuals living with hypertension (DREAM-GLOBAL): mixed methods study. JMIR Mhealth Uhealth 2019 Dec 09;7(12):e12639 [FREE Full text] [doi: 10.2196/12639] [Medline: 31815678]
46. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile apps for blood pressure monitoring: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
47. Cousins N. Anatomy of an illness (as perceived by the patient). N Engl J Med 1976 Dec 23;295(26):1458-1463. [doi: 10.1056/NEJM197612232952605] [Medline: 995143]
48. Knight EL, Bohn RL, Wang PS, Glynn RJ, Mogun H, Avorn J. Predictors of uncontrolled hypertension in ambulatory patients. Hypertension 2001 Oct;38(4):809-814. [doi: 10.1161/hy0901.091681] [Medline: 11641291]
49. Bokhour BG, Cohn ES, Cortés DE, Solomon JL, Fix GM, Elwy AR, et al. The role of patients' explanatory models and daily-lived experience in hypertension self-management. J Gen Intern Med 2012 Dec;27(12):1626-1634 [FREE Full text] [doi: 10.1007/s11606-012-2141-2] [Medline: 22821569]
50. Paterick TE, Patel N, Tajik AJ, Chandrasekaran K. Improving health outcomes through patient education and partnerships with patients. Proc (Bayl Univ Med Cent) 2017 Jan;30(1):112-113 [FREE Full text] [doi: 10.1080/08998280.2017.11929552] [Medline: 28152110]
51. Salari R, Kalhori SR, Fatehi F, Ghazisaeedi M, Nazari M. Determining minimum set of features for diabetes mobile apps. J Diabetes Metab Disord 2019 Dec;18(2):333-340 [FREE Full text] [doi: 10.1007/s40200-019-00417-y] [Medline: 31890658]
52. 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] [Medline: 30131313]
53. Garner SL, George CE, Young P, Hitchcock J, Koch H, Green G, et al. Effectiveness of an mHealth application to improve hypertension health literacy in India. Int Nurs Rev 2020 Dec;67(4):476-483. [doi: 10.1111/inr.12616] [Medline: 32767464]
54. Zarea V, Vishwa MV, Sattari M, Aliparasti R. High school students' perspective on the features of consumer health information websites. Webology 2016;13(1):1-14 [FREE Full text]

55. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
56. Rogers R, Prentice-Dunn S. Protection motivation theory. In: Gochman DS, editor. Handbook of Health Behavior Research I: Personal and Social Determinants. Switzerland: Springer; 1997:113-132.
57. Tsang JY, Blakeman T, Hegarty J, Humphreys J, Harvey G. Understanding the implementation of interventions to improve the management of chronic kidney disease in primary care: a rapid realist review. Implement Sci 2016 Apr 04;11:47 [FREE Full text] [doi: 10.1186/s13012-016-0413-7] [Medline: 27044401]
58. Dickson A, Knussen C, Flowers P. 'That was my old life; it's almost like a past-life now': identity crisis, loss and adjustment amongst people living with Chronic Fatigue Syndrome. Psychol Health 2008;23(4):459-476. [doi: 10.1080/08870440701757393] [Medline: 25160579]
59. Song T, Qian S, Yu P. Mobile health interventions for self-control of unhealthy alcohol use: systematic review. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e10899 [FREE Full text] [doi: 10.2196/10899] [Medline: 30694200]
60. Barley E, Lawson V. Using health psychology to help patients: theories of behaviour change. Br J Nurs 2016 Sep 08;25(16):924-927. [doi: 10.12968/bjon.2016.25.16.924] [Medline: 27615529]
61. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
62. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]

Abbreviations

BP: blood pressure
COM-B: Capability, Opportunity, Motivation, Behavior
mHealth: mobile health

Edited by R Kukafka; submitted 05.11.20; peer-reviewed by M Hawley, F Fischer; comments to author 15.12.20; revised version received 12.01.21; accepted 27.04.21; published 21.06.21.

Please cite as:
Song T, Liu F, Deng N, Qian S, Cui T, Guan Y, Arnolda L, Zhang Z, Yu P
A Comprehensive 6A Framework for Improving Patient Self-Management of Hypertension Using mHealth Services: Qualitative Thematic Analysis
J Med Internet Res 2021;23(6):e25522
URL: https://www.jmir.org/2021/6/e25522
doi:10.2196/25522
PMID:34152272

Original Paper

Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (Youper): Longitudinal Observational Study

Ashish Mehta[1], BA; Andrea Nicole Niles[2], PhD; Jose Hamilton Vargas[2], MD; Thiago Marafon[2], MSc; Diego Dotta Couto[2], MSc; James Jonathan Gross[1], PhD

[1]Department of Psychology, Stanford University, Stanford, CA, United States
[2]Youper, Inc, San Francisco, CA, United States

Corresponding Author:
Ashish Mehta, BA
Department of Psychology, Stanford University
Building 420
450 Jane Stanford Way
Stanford, CA, 94305
United States
Phone: 1 650 724 5436
Email: ashm@stanford.edu

Abstract

Background: Youper is a widely used, commercially available mobile app that uses artificial intelligence therapy for the treatment of anxiety and depression.

Objective: Our study examined the acceptability and effectiveness of Youper. Further, we tested the cumulative regulation hypothesis, which posits that cumulative emotion regulation successes with repeated intervention engagement will predict longer-term anxiety and depression symptom reduction.

Methods: We examined data from paying Youper users (N=4517) who allowed their data to be used for research. To characterize the acceptability of Youper, we asked users to rate the app on a 5-star scale and measured retention statistics for users' first 4 weeks of subscription. To examine effectiveness, we examined longitudinal measures of anxiety and depression symptoms. To test the cumulative regulation hypothesis, we used the proportion of successful emotion regulation attempts to predict symptom reduction.

Results: Youper users rated the app highly (mean 4.36 stars, SD 0.84), and 42.66% (1927/4517) of users were retained by week 4. Symptoms decreased in the first 2 weeks of app use (anxiety: d=0.57; depression: d=0.46). Anxiety improvements were maintained in the subsequent 2 weeks, but depression symptoms increased slightly with a very small effect size (d=0.05). A higher proportion of successful emotion regulation attempts significantly predicted greater anxiety and depression symptom reduction.

Conclusions: Youper is a low-cost, completely self-guided treatment that is accessible to users who may not otherwise access mental health care. Our findings demonstrate the acceptability and effectiveness of Youper as a treatment for anxiety and depression symptoms and support continued study of Youper in a randomized clinical trial.

(J Med Internet Res 2021;23(6):e26771) doi:10.2196/26771

KEYWORDS

digital mental health treatment; acceptability; effectiveness; anxiety; depression

Introduction

Nearly half the people in the United States will have a mental disorder at some point during their life span [1,2], and many more will have subthreshold symptoms. The most frequent mental health conditions are anxiety and depression, jointly termed "emotional disorders," and these impact 34% and 21% of people in the United States, respectively [3]. Despite the availability of effective treatments for emotional disorders, most people in need of treatment will not receive it [4]. Researchers have found that both structural (eg, financial, availability) and attitudinal barriers (eg, desire to handle problems independently) prevent patients from seeking mental health treatment [5]. Fully automated mental health intervention apps offer the promise of overcoming these barriers. By obviating the need for a trained clinician, the cost of treatment can be reduced by orders of magnitude and can be delivered to anyone with access to the internet. Moreover, fully automated treatments provide a means to treat patients who are uncomfortable seeking help from another person.

One particularly promising type of digital mental health intervention is a mobile app that can be installed on a person's mobile device. The Apple and Google Play stores organize these apps, centralizing the location where interventions can be accessed and allowing users to vet apps by reviewing their descriptions in the store and reading user reviews. Once installed on a person's phone, the mobile app medium makes it possible for people to access interventions anytime and anywhere. This opportunity has not been overlooked. Recent figures tally over 10,000 mental health apps available to consumers [6]. However, while the options for mental health treatment apps are at an all-time high, there is little research on the acceptability of available app-based treatments and whether they can actually reduce symptoms of psychopathology. In particular, mobile apps that are completely self-guided, and hence maximally accessible and scalable, are especially understudied. That said, a significant portion of the literature supports the efficacy of self-guided cognitive behavioral therapy administered via the internet, which patients may or may not be able to access on mobile phones depending on the intervention [7,8]. Thus, the potential for mobile apps to demonstrate similar efficacy is promising. The present study aimed to assess the acceptability and effectiveness of a self-guided intervention app called Youper that targets emotional disorders.

Although few in number, a handful of randomized controlled trials (RCTs) examining fully self-guided mental health intervention apps have shown promise. One mobile cognitive behavioral therapy intervention similar to Youper that employs a humanlike, chatbot interface found that college students experienced significantly reduced depression symptoms over the course of 2 weeks [9]. A number of RCTs have demonstrated the efficacy of self-guided mobile treatment programs for depression, including 1 testing problem-solving therapy and cognitive training [10], 3 testing cognitive behavioral therapy interventions [11-13], and 1 testing acceptance-based therapy [14]. Small-scale observational studies have also shown positive results for self-guided app-based treatments for symptoms of depression and anxiety [15,16].

Although these studies are promising, the evidence on self-guided digital mental health treatment is limited by small sample sizes obtained almost exclusively via RCTs. RCTs are considered the gold standard of evidence and play an indispensable role in assessing the efficacy of medical interventions. However, real-world evidence provides a necessary complement to understanding the impact of an intervention in the context that it will be received [17]. This fact has been acknowledged by numerous government bodies including the National Institute of Health, the Food and Drug Administration, and the European Medicines Agency, who have called for real-world evidence on interventions [18-20]. This call arises from the recognition of external validity shortcomings in RCTs due to factors such as inclusion criteria and differences in treatment adherence [21-23]

First, RCTs may be composed of different populations than those that naturally seek digital mental health treatment. Participants in RCTs are acquired through recruiting efforts and are then selected based on specific inclusion criteria. For example, participants in RCTs may be required to meet a minimum threshold of depression symptoms [10]. In the real world, digital mental health treatment (often distributed via smartphone app stores) is available to anyone with a smartphone and users span the range of depression symptomatology levels. Thus, the distribution of potential app users may or may not match the distribution of participants seen in the small body of existing literature on self-guided digital mental health. Further, it is plausible that participants who enroll in clinical mental health trials are more comfortable with seeking external mental health care than are users that discreetly download an app on their phone. Since targeting populations who carry stigma against seeking mental health care is an important goal and potential advantage of this technology, it is important that the populations being studied have equivalent attitudes to the population of potential users.

Second, treatment adherence in RCTs may systematically differ from real-world app usage because of the different experience that a participant in a clinical trial has compared to a user who downloads an app. In clinical trials, participants are often paid money to participate. Paid participants may feel a social obligation to adhere to the treatment plan. However, in real life, where app-based treatment is a completely individual experience, it is unclear whether adherence will be equivalent. Moreover, RCTs require, at minimum, an initial contact with study coordinators and sometimes additional contacts throughout the study. As contact with a treatment provider is known to increase adherence, this initial contact could boost levels of engagement [24]. Because the degree to which one engages and adheres to a treatment is related to treatment success [7], it is critical that we supplement evidence gained from RCTs with an understanding of how treatment recipients organically experience the app in the real world.

We define artificial intelligence (AI) therapy as a digital and fully automated, mobile, psychological treatment program that uses a conversational interface to deliver just-in-time adaptive interventions. The 3 key features that set AI therapy apart from traditional digital intervention approaches are (1) the use of a conversational (chatbot) interface, (2) inclusion of just-in-time interventions, and (3) adaptation and personalization. A primary goal of this study is to test whether AI therapy has potential as a viable treatment approach.

An additional goal of this study is to test the theoretical model underlying the just-in-time approach, which is a critical feature of AI therapy. Although just-in-time approaches have been used to target health behaviors, such as alcohol use, smoking, and obesity, only a small number of studies have described this approach in relation to emotional disorders [25,26].

Just-in-time interventions are designed to help the user manage moment-to-moment challenges that accumulate to negatively impact broader mental health functioning and produce symptoms of psychiatric disorders. Thus, just-in-time interventions target a proximal outcome that is theorized to accumulate over time to impact a longer-term outcome. In the case of anxiety and depression symptoms, emotion dysregulation is theorized to be a proximal cause for the manifestation of symptoms as well as a target for intervention [27,28]. Consistent with this hypothesis, a prior study of a just-in-time intervention for depression showed that it had promising impacts on depressive symptoms, albeit in a sample with just 10 people [25]. Following this theoretical work, we hypothesized that users who repeatedly succeed in regulating negative emotions by engaging just-in-time digital mental health interventions will experience long-term symptom reduction through accumulation of these regulation successes. We call this the *cumulative regulation hypothesis*. In addition to assessing AI therapy's effectiveness for symptom reduction, the current study will test the cumulative regulation hypothesis by testing the association between the accumulation of successful emotion regulation efforts with symptom reduction over time.

In this paper, we examine the acceptability and effectiveness of AI therapy as it is implemented in a smartphone app called Youper. Our study had 3 aims. First, we explored the acceptability of Youper by analyzing user ratings and retention metrics. Second, we examined effectiveness by measuring the reduction in anxiety and depression symptoms in the first month of app use. We hypothesized that users of Youper would experience a reduction in anxiety and depression symptoms during this time period. Third, we tested the cumulative regulation hypothesis by examining the longitudinal relationship between success in downregulating acute negative emotion in Youper conversations and clinical symptoms. We hypothesized that within-session emotion regulation success during Youper engagements would predict greater reductions in anxiety and depression symptoms. Finally, in exploratory analyses, we examined whether demographics, including gender and age; and clinical characteristics, including the number of self-reported diagnoses, current psychotropic medications, and concurrent therapy, could predict symptom reduction over time.

Methods

Participants

Participants were Youper subscribers (ie, users who paid for full access to Youper) who downloaded the app between March 4, 2020, and July 10, 2020. This time frame was selected because Youper was relatively stable during this period (ie, no significant updates or changes to the intervention were deployed during this time). Subscribers paid US $44.99 to have unlimited access to Youper's interventions for 1 year. Users who did not subscribe were only able to access the emotion regulation interventions once as part of a free sample, and therefore, were not included in this analysis.

Of 5943 users who completed at least one symptom measure in the study timeframe, 76.01% (n=4517) agreed for their data to be used for research, leaving a useable sample of 4517 participants. The sample was composed of 81.62% women (n=3687), 14.15% men (n=639), and 3.43% nonbinary individuals (n=155), and the average age of participants was 28.73 years (SD 9.63). Additional participant demographics and clinical characteristics are presented in Table 1. Participants completed symptom assessments at baseline (T0; within 3 days of subscribing to Youper), 2 weeks after baseline (T1), and 4 weeks after baseline (T2). Assessments were available to users every 14 days, and the majority of users completed their assessments within 3 days of them becoming available. Participants received access to anxiety or depression symptom measures based on their responses to screening questions. Symptom measures were administered if they endorsed a history of being diagnosed with clinical anxiety or depression, or if they reported elevated anxiety or depression symptoms on 2-item screening measures. Participants could receive only an anxiety measure, only a depression measure, or both depending on their responses to the screening items. Throughout the course of the measurement period, participants engaged in emotion regulation interventions at their discretion when emotional episodes arose.

Table 1. Additional demographic and clinical characteristics (N=4517).

Variable	Value, n (%)
Occupation	
Employed full time	2221 (49.17%)
Work and attend school	708 (15.67%)
Student full time	493 (10.91%)
Unemployed	420 (9.30%)
Freelancer or work part time	397 (8.79%)
Homemaker	149 (3.30%)
No response	93 (2.06%)
Retired	36 (0.80%)
Operating system	
iOS	4038 (89.40%)
Android	479 (10.60%)
Talked to a doctor about emotional health	
Yes	3559 (78.79%)
No	954 (21.12%)
Self-reported diagnoses[a,b]	
Anxiety disorder	2479 (54.88%)
Depressive disorder	2412 (53.40%)
Any diagnosis	2994 (66.28%)
Current treatment type[c]	
Prescribed medication	1904 (42.15%)
Psychotherapy	1196 (26.48%)
Prescribed medication or psychotherapy	2161 (47.84 %)

[a] Users were only asked about diagnoses if they reported talking to a doctor about their emotional health (n=3559).

[b] Mean number of diagnoses=2.82 (SD 1.47).

[c] Users were only asked about treatment if they reported a diagnosis (n=2994).

Youper Intervention

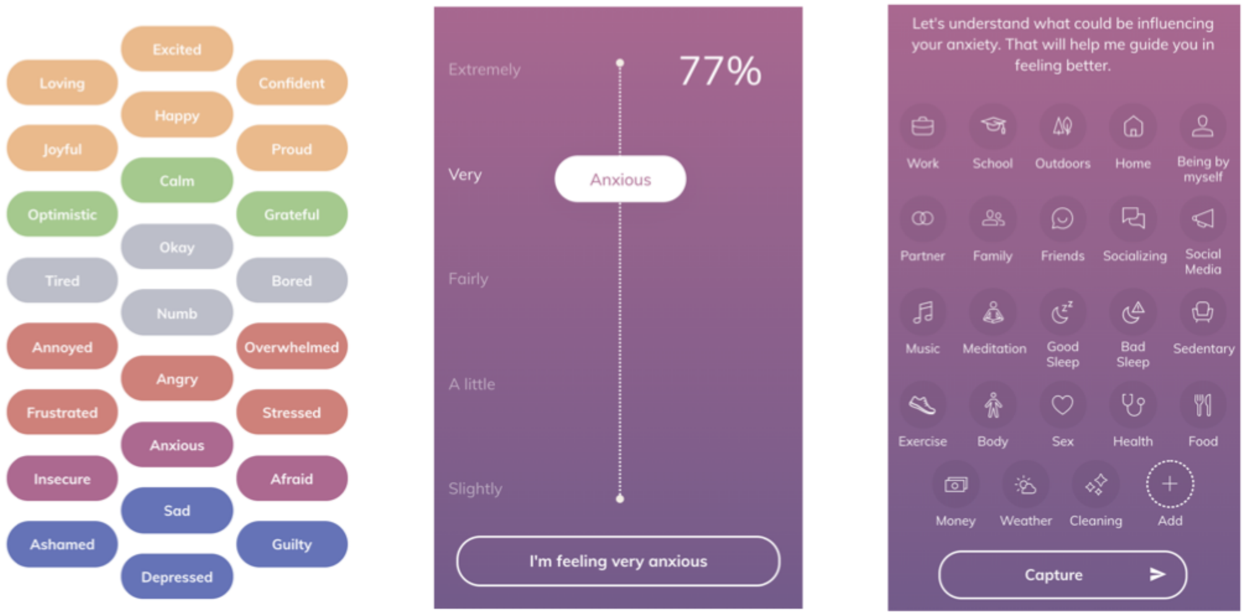
Youper is a novel intervention approach that aims to enhance the user's emotion regulation skills using empirically supported treatments for anxiety and depression. Although the emotion regulation strategies employed in Youper have precedent in existing treatment protocols for anxiety and depression, the adaptation of these interventions to help a user manage emotional distress at the present moment is novel. Youper's intervention is delivered via a conversational (ie, chat) interface and is entirely automated. Youper primarily uses a decision tree to select its responses to the user input. Each interaction with Youper is called a "conversation." Conversations follow a prespecified sequence (see Figure 1 for examples): identify current emotion and intensity (0%-100%), select contributing factors from a prespecified list, complete an open text entry about what is causing the current mood, complete emotion regulation skill practice for a negative mood or wellness practice for a positive mood (see Table 2), and identify current emotion and intensity (0%-100%).

The goal of each conversation is to help the user learn adaptive emotion regulation skills. If the user is experiencing a negative emotion, the skill targets the current emotion. If the user is experiencing a positive emotion, the skill encourages upregulation of that emotional state. If the user is in a neutral state, the skills encourage practice of activities that promote emotionally adaptive behaviors and attentional and cognitive control. Youper primarily uses just-in-time interventions (delivered at the moment of need) to help users practice and learn skills for emotion regulation. Youper's interventions target the 3 categories of treatment mechanisms defined by the common elements framework [29].

The common elements framework provides a review of common elements across cognitive and behavioral therapies inclusive of both traditional (cognitive therapy, behavioral activation) and third-wave (acceptance and commitment therapy, dialectical behavior therapy) approaches. They identify 3 mechanistic targets common to multiple effective therapies, including attention change (improving attentional focus and flexibility), cognitive change (improving ability to change perspective on an event), and context engagement (engaging new internal and external contexts to counteract maladaptive patterns). Youper's interventions aim to increase emotion regulation skills by targeting these common elements. For example, Youper includes interventions to increase attentional control such as mindfulness, cognitive change such as cognitive restructuring or gratitude journaling, and context engagement via behavioral activation exercises. The common elements framework was used to guide the development of Youper's interventions due to the extensive empirical support for the efficacy of each of these targets in enhancing emotion regulation and reducing symptoms of emotional disorders [30-36].

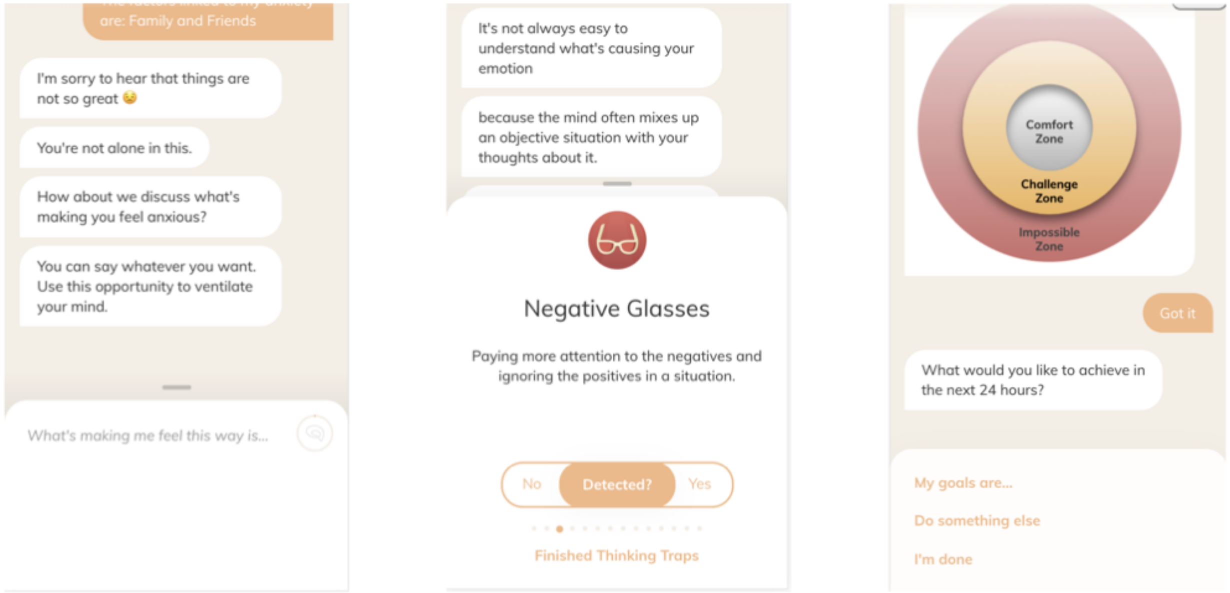
Each intervention is described in Table 2. Each skill follows a series of steps modeled after existing treatment manuals or research protocols. Skills practice includes a variety of formats including open-text entry following a prompt, graphical user interfaces, written content delivered via the chat, and audio.

Figure 1. Example interaction with Youper. Users start by reporting a discrete emotion (A) and the intensity (B) which they feel the emotion. They then report which factors contributed to the emotion (C) and describe the precipitating event (D). Next, they proceed through a randomly selected intervention (eg, E or F) from the list (see Table 2). Finally, they report their discrete emotional state again and the intensity which they feel that emotion (A and B).



Let's understand what could be influencing your anxiety. That will help me guide you in feeling better.

(A) Select discrete emotion (B) Select intensity (C) Select contributing factors



are: Family and Friends

I'm sorry to hear that things are not so great 😔

You're not alone in this.

How about we discuss what's making you feel anxious?

You can say whatever you want. Use this opportunity to ventilate your mind.

What's making me feel this way is...

It's not always easy to understand what's causing your emotion

because the mind often mixes up an objective situation with your thoughts about it.

Paying more attention to the negatives and ignoring the positives in a situation.

What would you like to achieve in the next 24 hours?

(D) Identify cause of emotion (E) Emotion regulation skill practice (eg, cognitive restructuring) (F) Emotion regulation skill practice (eg, goal setting)

Table 2. Emotion regulation change process targets and interventions.

Change process with interventions	Description
Context engagement	
Behavioral activation	1. Selecting a rewarding or social activity from a list 2. Psychoeducation about action/motivation cycle 3. Setting a goal to complete the activity
Goal setting	1. Psychoeducation about setting challenging, specific goals 2. Identifying a goal 3. Setting a reminder to check about goal completion
Problem solving	1. Identifying the problem 2. Identifying a goal 3. Brainstorming solutions 4. Selecting a solution and setting a goal
Attention change	
Mindfulness	1. Selection from a list of audio-recorded mindfulness exercises such as following the breath, progressive muscle relaxation, and mindfulness of thoughts
Sleep relaxation	1. Visualization of calming scenery 2. Selection from a list of different types of white noise
Cognitive change	
Acceptance	1. Practicing accepting negative thoughts and feelings without trying to change them 2. Planning to engage in value-driven behavior
Cognitive restructuring	1. Identifying thoughts 2. Identifying cognitive distortions 3. Examining evidence 4. Identifying alternative thoughts
Gratitude journaling	1. Identification of things for which the user is grateful
Self-compassion	1. Identification of how the user would treat a friend dealing with difficult emotions 2. Identification of how to treat oneself with the same compassion

Measures

Acceptability Measures

User Ratings

To assess acceptability of the Youper intervention, we asked users to provide a rating of the app using a 5-star scale. Users were given the following prompt: “I’d love to know how our journey together is going so far.” Users then provided their rating of Youper by selecting a number of stars ranging from 1 to 5. Users then were asked to provide feedback using an open text box.

Retention

Retention was measured as the proportion of Youper subscribers who engaged with the app during week 1, 2, 3, and 4 after subscription, as well as the average number of conversations that users had during each of these weeks.

Anxiety and Depression Symptoms

Anxiety symptoms were measured using the 7-item generalized anxiety disorder measure (GAD-7) [37]. The GAD-7 is a widely used measure of generalized anxiety disorder symptom severity and is frequently used as a general measure of overall anxiety symptoms. It has demonstrated excellent psychometrics with a Cronbach α of .92 and a sensitivity and specificity of 89% and 82%, respectively, for classifying generalized anxiety [37]. Depression symptoms were measured using a modified version of the Patient Health Questionnaire-9 (PHQ-9) with the suicide-related item removed and 1 of the items divided into 2 separate items [38]. Specifically, the item that asks if the respondent “has been moving slowly or has been fidgety and restless” was divided into an item about “moving slowly” and another item about “being fidgety and restless.” The PHQ-9 is a widely used measure of depression symptom severity with excellent psychometrics. The PHQ-9 has a Cronbach α of .89 and has both a sensitivity and a specificity of 88% for classifying major depression. In our slightly modified PHQ, we observed a comparable Cronbach α of .84 (95% CI 0.83-0.85) indicating good reliability.

Predictors of Symptom Reduction

Within-Session Emotion Regulation

To test the cumulative regulation hypothesis, we derived a measure of cumulative emotion regulation success. At the beginning of each Youper conversation, users selected their current emotion from a list of possible emotions as well as the intensity of that emotion (see Figure 1). Users who selected a negative emotion were also asked to report their emotion at the end of the conversation with Youper. We classified cases where users started with a negative emotion and ended with either a positive emotion or with a less intense negative emotion as a within-session emotion regulation "success." We classified cases where users reported a worsening or unchanging negative emotion as a "failure to regulate." To calculate a measure of cumulative within-session regulation success, we computed the proportion of cases classified as a success out of all conversations that started with a negative emotion.

As discrete negative emotion words encode different emotional intensities, we scaled the numeric self-reported emotional intensity according to an intensity scale factor corresponding to the discrete emotion the user selected. To derive the intensity scale factor for each discrete emotion, we first obtained normative valence and arousal ratings from a database of words that have been rated on a scale of 1 to 9 by a large sample of participants [39]. Next, we subtracted a constant (C=6) from the normative valence ratings, chosen so that all negative valence words would have negative-valued ratings and positive valence words would have positive ratings. To compute the intensity scale factor for each emotion word, we took the square root of the sum of the squared valence and arousal ratings (ie, the L2 norm). This decision was premised on the assumption that emotion intensity is a composite of valence and arousal [40]. Finally, we multiplied the self-reported numeric intensity by the intensity scale factor for the given emotion to obtain a scaled emotion intensity rating that could be compared across discrete emotion categories.

The scaling procedure had the effect of incorporating both the intensity of the emotion word and the self-reported numeric intensity into a single value which could be used to assess emotion regulation success pre- to postintervention. For example, without scaling, a participant that went from a rating of "75 annoyed" to "70 angry" would be erroneously classified as an instance of successful downregulation of negative emotion, despite the higher intensity imbued in the word "angry." With scaling, "75 annoyed" would translate to "–383" and "70 angry" would translate to "–482," and the increase in magnitude of negative emotion would result in a classification of failure to regulate. However, if the participant went from "75 annoyed" to "30 angry," the "30 angry" rating would be scaled to "–207," and the instance would be classified as a regulatory success. This procedure allowed us to use both the text information and numeric information in our assessment of success or failure to regulate emotions. As a check of robustness, we ran all analyses without the scaling procedure, and the results were substantively similar. To be conservative, we ultimately dichotomized these scaled scores into regulation successes and failures because, despite appearing to have a continuous measure of emotion regulation success, we were not confident that these scores truly represented precise gradations along a continuum.

Demographic and Clinical Characteristics

We examined both demographic and clinical characteristics as predictors of symptom reduction. Demographic characteristics included age (continuous) and gender (multinomial; man, woman, and nonbinary). Clinical characteristics included number of self-reported diagnoses (continuous), whether the user was currently taking psychotropic medication (binary), and whether the user was currently receiving psychotherapy (binary).

Statistical Analyses

Aim 1: Acceptability

We report descriptive statistics for user retention and app ratings.

Aim 2: Effectiveness

To estimate symptom reduction as a function of time, we fit piecewise multilevel models in R (version 4.0.2; The R Foundation for Statistical Computing) using the package "lmerTest" (version 3.1-2) [41]. Consistent with prior work, we selected a piecewise approach to capture a typical pattern of symptom reduction observed in treatment studies where symptoms initially decrease sharply and then level out as time progresses [42-47]. In these models, we regressed the symptom outcome measure (GAD-7 score or PHQ score) onto the number of days since subscribing to the app.

We selected multilevel models because our outcome measures were nested within individuals as a result of repeated measurement at multiple timepoints. Multilevel models allow for the estimation of within-subject effects. Further, when fit with maximum likelihood, multilevel models allow for the inclusion of participants with incomplete data without deletion or imputation and produces unbiased estimates for model parameters [48,49]. As per guidelines for randomized clinical trials, we conducted an intent-to-treat analysis, including all participants who had at least one assessment [50-53]. As discussed by Gupta [51], "intent-to-treat analysis avoids overoptimistic estimates of the efficacy of an intervention resulting from the removal of non-compliers by accepting that noncompliance and protocol deviations are likely to occur in actual clinical practice." We estimated the reduction of symptoms from T0 to T1 and from T1 to T2. We used a breakpoint at 14 days, as participants' second of 3 symptom measurements was available to be completed 14 days after the first measurement. Because not all participants completed assessments immediately when they were available, we chose to treat time as a continuous predictor in our analysis rather than simply grouping observations into time points at T0, T1, and T2. This approach allowed us to keep all information that we had about the time that had elapsed from baseline and was more conservative because it did not assume that a change occurring more than 14 days after baseline was occurring exactly at 14 days.

The models included 2 fixed effect parameters: one which estimated the slope of symptom reduction from the start of using Youper to 14 days later, and another which estimated the slope of symptom reduction from the 14-day mark onward.

Additionally, we included a random intercept term for each participant. We calculated Cohen *d* effect sizes by dividing the mean difference in symptom levels by the square root of the sum of the participant-level intercept variance and the residual variance [54].

Aim 3: Cumulative Regulation Hypothesis

To test the cumulative regulation hypothesis (ie, whether cumulative emotion regulation success within conversations predicted subsequent psychopathology symptoms), we fit longitudinal path analysis models for each of the 2 symptom measures (GAD-7 and PHQ) in the R package, “lavaan” (version 0.6-6) [55]. We fit these models using full information maximum likelihood and allowed the covariances between exogenous variables to be freely estimated [56]. This method enabled us to conduct an intent-to-treat analysis, including all participants that had a measurement for at least one variable included in the model. In these models, we estimated all autoregressive paths and lagged paths from emotion regulation success to subsequent clinical symptoms. Specifically, each path analysis model consisted of 3 regression equations. In the first equation, we regressed the T1 symptom outcome onto the T0 symptom outcome and the proportion of within-session regulation successes between T0 and T1 (ie, the proportion of negative emotions that were successfully regulated of the total number of negative emotion regulation attempts). In the second equation, we regressed the proportion of within-session regulation successes between T1 and T2 onto the proportion of within-session regulation successes between T0 and T1. Finally, we regressed the T2 symptom outcome onto the T1 symptom outcome and the proportion of within-session regulation successes between T1 and T2. (See Figure 3 for an illustration of paths with standardized coefficients.)

Exploratory Analyses: Clinical and Demographic Predictors

To test predictors of treatment response, we fit piecewise mixed effects models like those used in Aim 2, with the addition of interaction terms for the specified predictor. Specifically, we regressed the symptom outcome onto the interaction of the specified predictor and the number of days since the participant subscribed to the app. We examined age (continuous), gender (dummy coded with female as the reference group), number of self-reported diagnoses (continuous), whether the user was taking psychotropic medication (binary), and a whether the user was in therapy (binary) as individual difference predictors of symptom reduction.

Results

Aim 1: Acceptability

User Ratings

On the 5-point star rating scale, the median Youper rating was 5 and the mean was 4.36 (SD 0.84). Out of 3667 users who rated the app, 56.09% (n=2057) gave a 5-star rating, 26.89% (n=986) gave 4 stars, 14.97% (n=549) gave 3 stars, 1.31% (n=48) gave 2 stars, and 0.74% (n=27) gave 1 star. The mean conversation number at which users provided a rating was conversation 1.40, and the SD was 4.03.

Retention

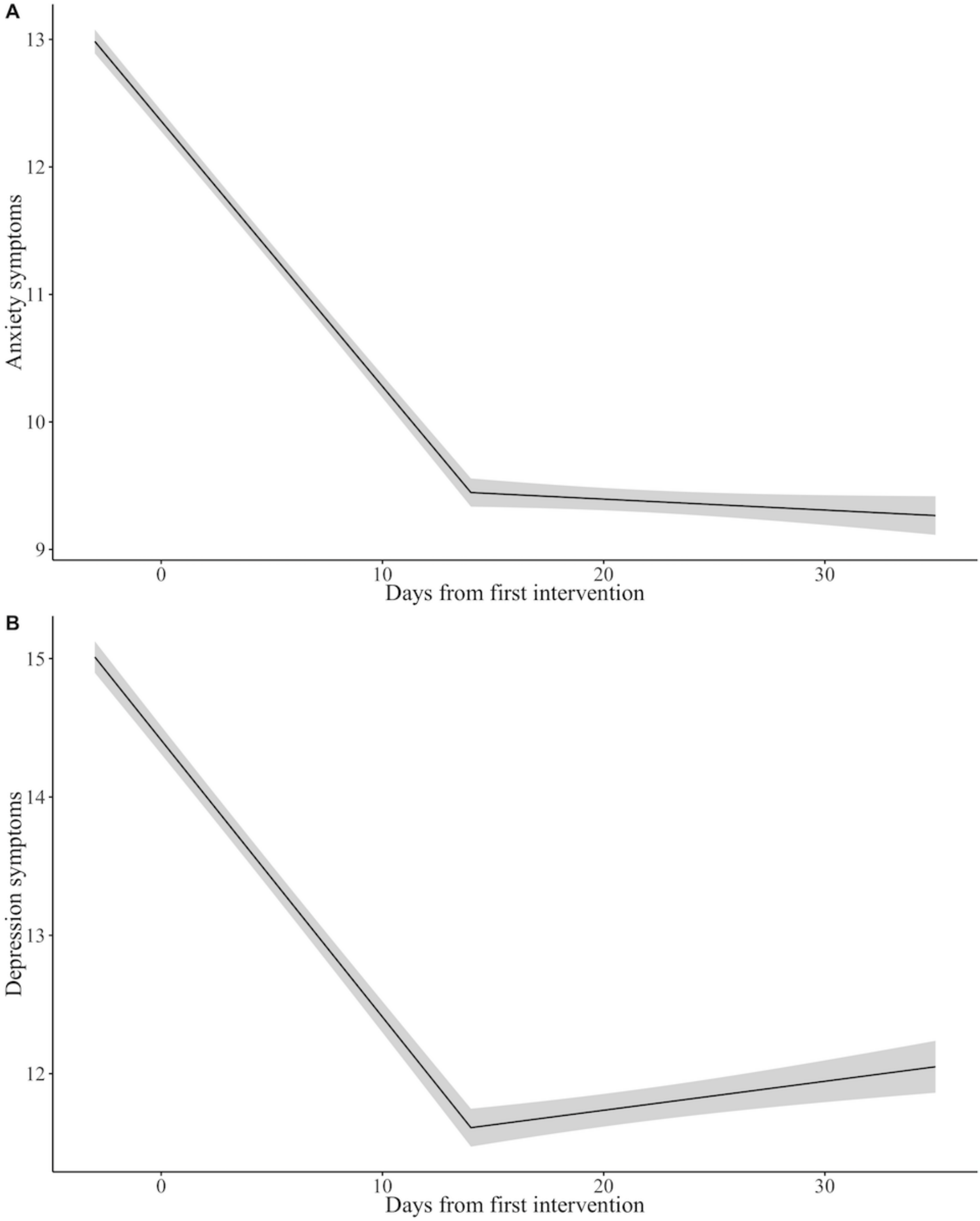
Of the 4517 users who subscribed to Youper between March 4, 2020, and July 10, 2020, 90.75% (n=4099) were still using Youper in week 1, 60.44% (n=2730) were using in week 2, 51.78% (n =2339) were using in week 3, and 42.66% (n=1927) were using in week 4 after subscription. The average numbers of conversations users had in weeks 1, 2, 3, and 4 were 6.50 (SD 6.74), 3.08 (SD 4.50), 2.36 (SD 3.98), and 2.04 (SD 3.83), respectively. Across the whole 4-week period, users engaged in an average of 13.98 conversations (SD 16.89).

Aim 2: Effectiveness

Anxiety

Results are displayed in Figure 2. Participants ($N_{participants}$=4144; $N_{observations}$=7093) experienced a significant reduction in anxiety symptoms from T0 to T1 (b=–0.21; bootstrapped 95% CI –0.22 to –0.19; P<.001). From T1 to T2, there was no significant change in anxiety symptoms (P=.35). The conditional means (and bootstrapped SEs) at day 0, day 14, and day 28 were 12.36 (SE 0.08), 9.45 (SE 0.11), and 9.33 (SE 0.11), respectively. These differences equate to Cohen *d*s of 0.57 between day 0 and day 14, 0.60 between day 0 and day 28, and 0.02 between day 14 and day 28. When analyses were conducted only on participants who had completed at least two assessments ($N_{participants}$=2117; $N_{observations}$=5066) or all 3 assessments ($N_{participants}$=827; $N_{observations}$=2481), results were unchanged.

Figure 2. Symptom reduction over time in the full sample. The gray shaded region indicates bootstrapped SEs. Model details are described in the Results for Aim 2.



Depression

Results are displayed in Figure 2. Participants ($N_{participants}$=3992; $N_{observations}$=6685) experienced a significant reduction in depression symptoms from T0 to T1 (*b*=–0.20; bootstrapped 95% CI –0.22 to –0.18; *P*<.001). From T1 to T2, depression symptoms increased slightly (*b* =0.02; bootstrapped 95% CI 0.00-0.04; *P*=.05). The conditional means (and bootstrapped SEs) at day 0, day 14, and day 28 were 14.41 (SE 0.10), 11.61 (SE 0.14), and 11.90 (SE 0.13), respectively. These differences equate to a Cohen *d* of 0.46 between day 0 and day 14, 0.42 between day 0 and day 28, and 0.05 between day 14 and day 28. When analyses were conducted only on participants who had completed at least two assessments ($N_{participants}$=1951; $N_{observations}$=4644) or all 3 assessments ($N_{participants}$=737; $N_{observations}$ = 2211), results were unchanged with the exception that symptoms no longer significantly increased between T1 and T2 (*P*=.08 and *P*=.15, respectively).

Aim 3: Cumulative Regulation Hypothesis

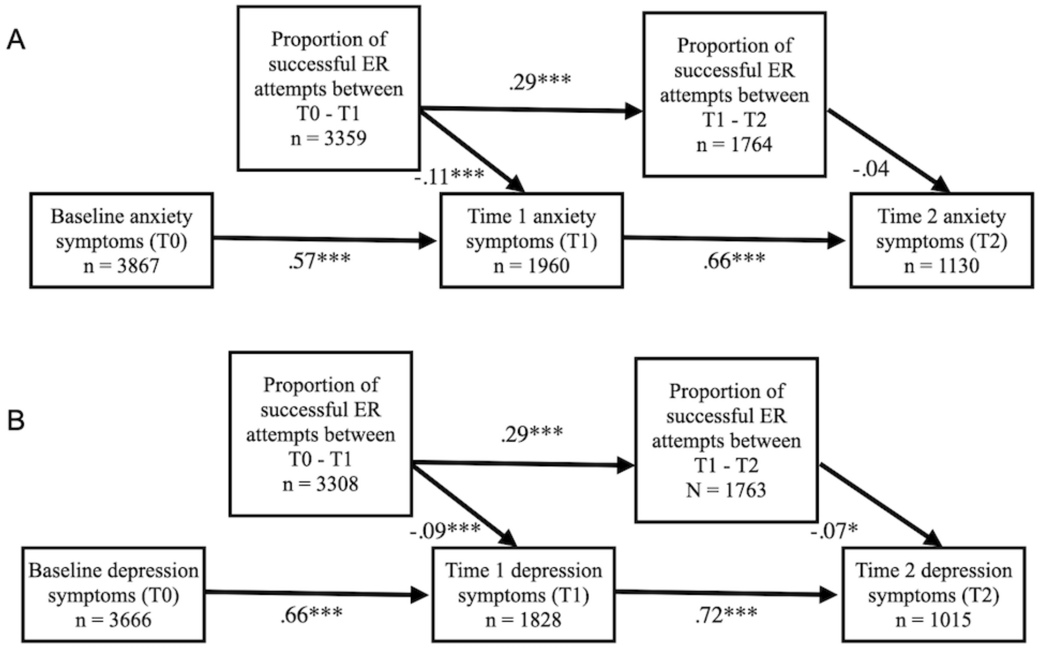
Preliminary Analyses

We first examined the probability that users would successfully regulate their emotion within a conversation with Youper. As described in the methods, we defined successful regulation as a conversation that started with a negative emotion and ended with either a negative emotion at a lower intensity or a positive emotion. Using a generalized linear model with logit link function and random intercepts for each participant and each preintervention discrete emotion ($N_{participants}$=4120; $N_{observations}$=32,885), we found that overall, participants were more likely to succeed in regulating their negative emotion than to fail (OR 4.82, bootstrapped 95% CI 3.89-5.99; *P*<.001).

Anxiety

To examine the effect of regulatory success within Youper sessions on anxiety symptoms, we fit a longitudinal path analysis model ($N_{participants}$=4284; see Figure 3 for ns for each variable). The model had good fit characteristics as indicated by a significant chi-square value and standard fit statistics (X^2_4=60.84; *P*<.001; root mean square of approximation [RMSEA]=0.058; Tucker-Lewis index [TLI]=0.91; comparative fit index [CFI]=0.96; standardized root mean squared residual [SRMR]=0.046). This model estimated that for each 0.10 increase in the proportion of negative emotions that users successfully improved between T0 and T1, users reported a 0.20 point reduction on the GAD-7 anxiety measure at T1 (*P*<.001). The effect of the proportion of negative emotions that users improved between T1 and T2 did not significantly reduce subsequent GAD-7 scores at T2 (*P*=.19). See Figure 3a for standardized coefficients for all paths.

Figure 3. Path analysis diagram with standardized coefficients. This diagram shows the autoregressive and lagged relationships between the proportion of a user’s ER attempts that were successful out of their total regulation attempts and subsequent anxiety symptoms (A) or depression symptoms (B). ER: emotion regulation. ****P*<.001, ***P*<.01, **P*<.05. Exact *P* values are noted in the text.



Depression

In order to examine the effect of emotion regulatory success on depression symptoms, we fit a similar longitudinal path analysis model ($N_{participants}$=4228; see Figure 3 for ns for each variable). This model also had good fit characteristics as indicated by a significant chi-square value and standard fit statistics (X^2_4=50.93; *P*<.001; RMSEA=0.053; TLI=0.94; CFI=0.97; SRMR=0.041). For each 0.10 increase in the proportion of negative emotions that users successfully regulated between T0 and T1, they reported a 0.20 point reduction on the subsequent PHQ depression measure at T1 (*P*<.001). For every 0.10 increase in the proportion of negative emotions that users successfully regulated between T1 and T2, they reported a 0.13 point reduction in depression symptoms at T2 (*P*=.02). See Figure 3b for standardized coefficients for all paths.

Secondary Analyses

In addition to our primary hypotheses, we also conducted exploratory analyses of potential individual difference predictors of symptom reduction. In these analyses, we fit piecewise mixed effects models with a breakpoint at 14 days (time of T1 symptom assessment). We regressed the specified symptom assessment onto the interaction of the specified predictor and the number of days since the user subscribed to the app. We examined age, gender, whether the user was taking psychotropic medication, and whether the user was in therapy as individual difference predictors of symptom reduction.

Age and Gender

The interaction effects of time using Youper with age on anxiety ($N_{participants}$=4143; $N_{observations}$=7090) and depression ($N_{participants}$=3991; $N_{observations}$=6683) symptoms were not significant from T0 to T1 ($P_{anxiety}$=.77; $P_{depression}$=.39) or from T1 to T2 ($P_{anxiety}$=.54; $P_{depression}$=.43) in the piecewise regression models.

The interaction effects of time using Youper with gender on anxiety ($N_{participants}$=4144; $N_{observations}$=7093) and depression ($N_{participants}$=3992; $N_{observations}$=6685) symptoms were not significant from T0 to T1 ($P_{anxiety-male}$=.74; $P_{anxiety-nonbinary}$=.44; $P_{anxiety-not-informed}$=.99; $P_{depression-male}$=.16; $P_{depression-nonbinary}$=.32; $P_{depression-not-informed}$=.37) or from T1 to T2 ($P_{anxiety-male}$=.40; $P_{anxiety-nonbinary}$=.28; $P_{anxiety-not-informed}$=.17; $P_{depression-male}$=.89; $P_{depression-non-binary}$=.70; $P_{depression-not-informed}$=.25) in the piecewise regression models.

Number of Self-reported Diagnoses

The interaction effects of the number of self-reported diagnoses with time using Youper on anxiety symptoms ($N_{participants}$=2679; $N_{observations}$=4661) was not significant from T1 to T2 (*P*=.08), and not significant from T0 to T1 (*P*=.59).

There was a significant interaction effect of number of self-reported diagnoses with time using Youper on depression symptoms from T1 to T2 ($N_{participants}$=2738; $N_{observations}$=4589; *b*=0.02; bootstrapped 95% CI 0.007-0.04; *P*=.006), but not from T0 to T1 (*P*=.78). This indicated that users with more diagnoses regressed modestly towards their baseline level of depression in the latter half of the treatment, whereas users with fewer diagnoses retained the treatment benefit.

Medication and Therapy

There were no significant interaction effects of taking prescribed medication with time using Youper on anxiety ($N_{participants}$=2719; $N_{observations}$=4733) or depression ($N_{participants}$=2776; $N_{observations}$=4654) symptoms from T0 to T1 ($P_{anxiety}$=.32.; $P_{depression}$=.72) or from T1 to T2 ($P_{anxiety}$=.57; $P_{depression}$=.66).

There were no significant interaction effects of receiving psychotherapy with time using Youper on anxiety ($N_{participants}$=2719; $N_{observations}$=4733) or depression ($N_{participants}$=2776; $N_{observations}$=4654) symptoms from T0 to T1 ($P_{anxiety}$=.66; $P_{depression}$=.87) or from T1 to T2 ($P_{anxiety}$=.65; $P_{depression}$=.52).

Discussion

Summary

The present study had 3 aims. First, we examined the acceptability of Youper AI therapy by assessing user ratings and retention metrics among subscribers. Second, we tested whether there were significant reductions in anxiety and depression symptoms. Third, we examined the cumulative regulation hypothesis, which predicts that the frequency of within-conversation emotion regulation success would predict symptom reduction.

Findings indicated that users were well retained and provided high ratings of Youper (median 5/5). As hypothesized, users showed significant reductions in symptoms in the first 2 weeks of using Youper with sustained improvements through 4 weeks from initial download. Finally, consistent with the cumulative regulation hypothesis, greater frequency of within-conversation emotion regulation successes significantly predicted greater reductions in anxiety and depression. Although no demographic predictors emerged, users with more self-reported diagnosed psychiatric conditions showed a slight return of depression symptoms between 2 and 4 weeks from first subscribing to Youper.

Acceptability and Effectiveness

Because retention poses a significant challenge for entirely unguided treatment programs, our finding that 60.44% (2730/4517) of users continued to engage with the app in the second week and 42.66% (1927/4517) of users continued to engage with the app in the fourth week after initial download is promising. Although there are no clearly established metrics of retention for mobile apps, a recent paper examining retention among different mobile apps showed that Youper had the highest “stickiness” (measured by the ratio of active users to downloads in a given month) compared to any other treatment app for anxiety and depression [57]. Because Youper users experienced symptom improvements on average within the first 2 weeks of app use, with the present retention rate, it is likely that a large portion of users will stick with the app long enough to experience some positive effects. It is also notable that the median satisfaction rating given by users was 5 out of 5. Taken together, these findings indicate that Youper has great potential as a highly acceptable and adequately engaging digital treatment program.

Youper users showed a moderate effect size reduction for anxiety (*d*=0.57) and depression (*d*=0.46) within 2 weeks of starting app use. The reduction in anxiety symptoms was maintained through the 4-week period (day 0 to day 28: *d*=0.60). The reduction in depression symptoms was maintained through the 4-week period (day 0 to day 28: *d*=0.42) although depression increased slightly, but significantly, between weeks 2 and 4. These effect sizes are comparable to those found in RCTs of other commercially available mobile apps tested for a similar duration [9-11,58], suggesting that the AI therapy approach is viable for further testing in a randomized clinical trial. Youper users also had high success at regulating their negative emotions with each conversation. Given the low cost and potential for broad dissemination of Youper, these findings are particularly exciting, as they provide preliminary evidence of Youper’s effectiveness as an emotion regulation tool and a transdiagnostic treatment. It is important to note, however, that the final mean PHQ score of 11.9 still fell in the moderate severity range. Thus, as we begin to understand the mechanisms of the AI therapy approach and gain greater understanding of how to maximize user engagement, we are hopeful that effects on symptom reduction will continue to improve.

Youper's symptom reduction, retention, and satisfaction ratings are notable because they were demonstrated in a real-world setting. Although highly controlled feasibility pilot trials allow determination of causal inference, these studies may not be generalizable to real-world settings and may fail to address issues of external relevance and dissemination [59]. Our analysis included a very large sample of Youper users who voluntarily downloaded and purchased the Youper program. Unlike in typical research settings, users were not recruited to participate or compensated for their assessments or for providing their feedback during their participation. Observed retention rates and symptom reduction therefore have already been shown in a real-world setting and population.

Cumulative Regulation Hypothesis

The finding that cumulative within-session emotion regulation was strongly predictive of symptom reduction provides preliminary evidence for a potential mechanism of the AI therapy just-in-time intervention approach. Youper is theorized to enact its effects by enhancing emotion regulation skills via -in-time interventions. Thus, more effective emotion regulation sessions would indicate progress towards enhanced general emotion regulation skills and ultimately, symptom reduction. Therefore, it is promising that the effectiveness of the emotion regulation practice predicts the longer-term impacts of app use on symptom reduction. Although these results provide initial support for the theorized model underlying Youper's treatment approach, randomization is critical for rigorously testing within-session emotion regulation as a mediator of symptom reduction.

Predictors of Symptom Reduction

Interestingly, no demographic predictors of symptom reduction emerged. These findings are largely consistent with the existing literature where demographic features rarely predict symptom reduction [60-67]. These findings are promising, suggesting that digital treatment programs can be broadly disseminated with similar potential benefit across demographic groups. The number of comorbid diagnoses was a significant predictor of response such that users who reported more diagnosed mental health conditions showed a slight return of depression symptoms between 2 and 4 weeks from the first subscription date. These findings are consistent with prior literature showing poorer outcomes with greater comorbidity in depression treatment [68-70]. Users with more diagnosed conditions likely have a more severe clinical presentation, meaning that an entirely self-guided program may be less effective for this group. The finding that concurrent medication and therapy did not significantly impact symptom reduction suggests that the demonstrated effects of Youper on symptoms are unlikely to be explained by concurrent treatment, and that participating in other treatments alongside Youper does not hinder its effects.

Limitations and Future Directions

Despite many strengths, our study had a few limitations. First, because these data were not collected as part of a research study, we did not have a control group, making it impossible to determine whether symptom reduction was simply due to the passage of time. However, given that effect sizes for symptom reduction that we found are comparable to those found in RCTs of other mobile app programs that showed significant differences between active treatment groups and wait list controls [9-11,58], it is unlikely that these effects can be explained by spontaneous remission. Second, because this was an observational study, we used the symptom data that were available to us, which included only self-report measures. Although we used validated measures, solely relying on self-report does not give a complete picture of the impact of Youper on clinical symptoms and overall functioning that could be more thoroughly assessed via clinical interviews. Third, this study included only 2 brief measures as outcomes: the PHQ and the GAD-7. Although these measures are widely used and show excellent psychometric properties, additional measures of anxiety, depression, and other purported outcome targets, such as quality of life and functioning, could help us better understand Youper's effectiveness. Fourth, 47.84% (2161/4517) of Youper users were concurrently taking medication or engaging in therapy, meaning that it is possible symptom reduction resulted from participation in these other treatments rather than Youper (although concurrent treatment was not a significant moderator of symptom reduction). Finally, our emotion regulation measure was not designed to assess the magnitude of emotion regulation success, meaning that our metric included only success or failure with each conversation. These limitations should be addressed in future studies that include a control group, that assess symptoms using clinician-administered measures, that include a broader array of self-report measures, and that use more precise measures of emotion regulation success.

Conclusions

This study provides preliminary evidence for Youper's acceptability in a real-world setting that is unfettered by the constraints of highly controlled clinical trials. It also provides evidence of Youper's effectiveness as an entirely unguided intervention for anxiety and depression. Finally, we demonstrated that Youper's effects on symptom reduction may be explained by repeated within-session emotion regulation successes, providing preliminary support for the process by which a just-in-time intervention can be effective for the treatment of emotional disorders. Our results highlight the potential impact of Youper as a low-cost, light-touch, transdiagnostic intervention for anxiety and depression that can be broadly disseminated to improve mental health for millions of people around the world.

Conflicts of Interest

Authors ANN, JHV, TM, and DDC are employees of Youper and shareholders in the company. AM and JJG have no conflicts of interest to declare.

References

1. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
2. World Health Organization. The World Health Report 2001: Mental Health: New Understanding, New Hope. Geneva, Switzerland: World Health Organization; 2001.
3. Kessler R, Petukhova M, Sampson N, Zaslavsky A, Wittchen HU. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
4. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
5. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
6. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 May 01;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
7. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: A meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
8. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
9. 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]
10. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
11. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
12. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of superbetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
13. Birney AJ, Gunn R, Russell JK, Ary DV. Moodhacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
14. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
15. 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]
16. Miller CB, Gu J, Henry AL, Davis ML, Espie CA, Stott R, et al. Feasibility and efficacy of a digital CBT intervention for symptoms of generalized anxiety disorder: a randomized multiple-baseline study. J Behav Ther Exp Psychiatry 2021 Mar;70:101609. [doi: 10.1016/j.jbtep.2020.101609] [Medline: 32950939]
17. Monti S, Grosso V, Todoerti M, Caporali R. Randomized controlled trials and real-world data: differences and similarities to untangle literature data. Rheumatology (Oxford) 2018 Oct 01;57(57 Suppl 7):vii54-vii58. [doi: 10.1093/rheumatology/key109] [Medline: 30289534]
18. Food and Drug Administration. Developing a framework for regulatory use of real-world evidence. govinfo. URL: https://www.govinfo.gov/content/pkg/FR-2017-07-31/pdf/2017-16021.pdf [accessed 2021-01-26]
19. Arlett P. Update on Real World Evidence Data Collection. European Medicines Agency. 2016 Mar. URL: https://ec.europa.eu/health/sites/health/files/files/committee/stamp/2016-03_stamp4/4_real_world_evidence_ema_presentation.pdf [accessed 2021-01-26]
20. HTA 101: Introduction to health technology assessment internet. U.S. National Library of Medicine. URL: https://www.nlm.nih.gov/nichsr/hta101/ta10103.html [accessed 2012-01-25]
21. Travers J, Marsh S, Williams M, Weatherall M, Caldwell B, Shirtcliffe P, et al. External validity of randomised controlled trials in asthma: to whom do the results of the trials apply? Thorax 2007 Mar 01;62(3):219-223 [FREE Full text] [doi: 10.1136/thx.2006.066837] [Medline: 17105779]

22. Fortin M, Dionne J, Pinho G, Gignac J, Almirall J, Lapointe L. Randomized controlled trials: do they have external validity for patients with multiple comorbidities? Ann Fam Med 2006;4(2):104-108 [FREE Full text] [doi: 10.1370/afm.516] [Medline: 16569712]
23. Carls GS, Tuttle E, Tan R, Huynh J, Yee J, Edelman SV, et al. Understanding the gap between efficacy in randomized controlled trials and effectiveness in real-world use of GLP-1 ra and DPP-4 therapies in patients with type 2 diabetes. Diabetes Care 2017 Nov 11;40(11):1469-1478. [doi: 10.2337/dc16-2725] [Medline: 28801475]
24. Alfonsson S, Olsson E, Linderman S, Winnerhed S, Hursti T. Is online treatment adherence affected by presentation and therapist support? A randomized controlled trial. Computers in Human Behavior 2016 Jul;60:550-558. [doi: 10.1016/j.chb.2016.01.035]
25. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
26. Ghandeharioun A, McDuff D, Czerwinski M, Rowan K. EMMA: an emotion-aware wellbeing chatbot. In: arXiv. 2019 Presented at: 8th International Conference on Affective Computing and Intelligent Interaction (ACII); 2019 Sept 3-6; Cambridge, UK p. 1-7 URL: http://arxiv.org/abs/1812.11423 [doi: 10.1109/acii.2019.8925455]
27. Hofmann S, Sawyer A, Fang A, Asnaani A. Emotion dysregulation model of mood and anxiety disorders. Depress Anxiety 2012 May;29(5):409-416 [FREE Full text] [doi: 10.1002/da.21888] [Medline: 22430982]
28. Sheppes G, Suri G, Gross JJ. Emotion regulation and psychopathology. Annu Rev Clin Psychol 2015 Mar 28;11(1):379-405. [doi: 10.1146/annurev-clinpsy-032814-112739] [Medline: 25581242]
29. Mennin DS, Ellard KK, Fresco DM, Gross JJ. United we stand: emphasizing commonalities across cognitive-behavioral therapies. Behav Ther 2013 Jun;44(2):234-248 [FREE Full text] [doi: 10.1016/j.beth.2013.02.004] [Medline: 23611074]
30. Barlow D, Farchione TJ, editors. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Therapist Guide (2nd ed.). New York: Oxford University Press; 2018.
31. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive Therapy of Depression. New York: Guilford Press; 1979.
32. Chambers R, Lo BCY, Allen NB. The impact of intensive mindfulness training on attentional control, cognitive style, and affect. Cogn Ther Res 2007 Feb 23;32(3):303-322. [doi: 10.1007/s10608-007-9119-0]
33. Craske MG, Kircanski K, Zelikowsky M, Mystkowski J, Chowdhury N, Baker A. Optimizing inhibitory learning during exposure therapy. Behav Res Ther 2008 Jan;46(1):5-27. [doi: 10.1016/j.brat.2007.10.003] [Medline: 18005936]
34. Forman EM, Herbert JD, Moitra E, Yeomans PD, Geller PA. A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. Behav Modif 2007 Nov 26;31(6):772-799. [doi: 10.1177/0145445507302202] [Medline: 17932235]
35. Lohr JM, Olatunji BO, Sawchuk CN. A functional analysis of danger and safety signals in anxiety disorders. Clin Psychol Rev 2007 Jan;27(1):114-126. [doi: 10.1016/j.cpr.2006.07.005] [Medline: 16997437]
36. Roemer L, Williston SK, Rollins LG. Mindfulness and emotion regulation. Current Opinion in Psychology 2015 Jun;3:52-57. [doi: 10.1016/j.copsyc.2015.02.006]
37. 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]
38. 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]
39. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav Res Methods 2013 Dec 13;45(4):1191-1207. [doi: 10.3758/s13428-012-0314-x] [Medline: 23404613]
40. Russell JA. A circumplex model of affect. Journal of Personality and Social Psychology 1980;39(6):1161-1178. [doi: 10.1037/h0077714]
41. Kuznetsova A, Brockhoff PB, Christensen RHB. Package: tests in linear mixed effects models. J. Stat. Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
42. Craske MG, Niles AN, Burklund LJ, Wolitzky-Taylor KB, Vilardaga JCP, Arch JJ, et al. Randomized controlled trial of cognitive behavioral therapy and acceptance and commitment therapy for social phobia: outcomes and moderators. J Consult Clin Psychol 2014 Dec;82(6):1034-1048 [FREE Full text] [doi: 10.1037/a0037212] [Medline: 24999670]
43. Roy-Byrne P, Craske MG, Sullivan G, Rose RD, Edlund MJ, Lang AJ, et al. Delivery of evidence-based treatment for multiple anxiety disorders in primary care: a randomized controlled trial. JAMA 2010 May 19;303(19):1921-1928 [FREE Full text] [doi: 10.1001/jama.2010.608] [Medline: 20483968]
44. Niles AN, Craske MG, Lieberman MD, Hur C. Affect labeling enhances exposure effectiveness for public speaking anxiety. Behav Res Ther 2015 May;68:27-36. [doi: 10.1016/j.brat.2015.03.004] [Medline: 25795524]
45. Wolitzky-Taylor K, Niles AN, Ries R, Krull JL, Rawson R, Roy-Byrne P, et al. Who needs more than standard care? Treatment moderators in a randomized clinical trial comparing addiction treatment alone to addiction treatment plus anxiety disorder treatment for comorbid anxiety and substance use disorders. Behav Res Ther 2018 Aug;107:1-9. [doi: 10.1016/j.brat.2018.05.005] [Medline: 29800622]

46. Niles AN, Axelsson E, Andersson E, Hedman-Lagerlöf E, Carlbring P, Andersson G, et al. Internet-based cognitive behavior therapy for depression, social anxiety disorder, and panic disorder: Effectiveness and predictors of response in a teaching clinic. Behav Res Ther 2021 Jan;136:103767. [doi: 10.1016/j.brat.2020.103767] [Medline: 33249272]
47. Niles AN, Woolley JD, Tripp P, Pesquita A, Vinogradov S, Neylan TC, et al. Randomized controlled trial testing mobile-based attention-bias modification for posttraumatic stress using personalized word stimuli. Clinical Psychological Science 2020 Jun 05;8(4):756-772. [doi: 10.1177/2167702620902119]
48. Snijders T, Bosker R. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. London: SAGE; 2012.
49. Little R, Rubin D. Statistical Analysis With Missing Data. New York: Wiley; 1987.
50. Detry MA, Lewis RJ. The intention-to-treat principle: how to assess the true effect of choosing a medical treatment. JAMA 2014 Jul 02;312(1):85-86. [doi: 10.1001/jama.2014.7523] [Medline: 25058221]
51. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
52. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999 Sep 11;319(7211):670-674 [FREE Full text] [doi: 10.1136/bmj.319.7211.670] [Medline: 10480822]
53. McCoy E. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017 Oct 18;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
54. Westfall J, Kenny DA, Judd CM. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. J Exp Psychol Gen 2014 Oct;143(5):2020-2045. [doi: 10.1037/xge0000014] [Medline: 25111580]
55. Rosseel Y. lavaan: an r package for structural equation modeling. J. Stat. Soft 2012;48(2):5-12. [doi: 10.18637/jss.v048.i02]
56. Allison PD. Missing data techniques for structural equation modeling. Journal of Abnormal Psychology 2003;112(4):545-557. [doi: 10.1037/0021-843x.112.4.545]
57. Carlo AD, Hosseini Ghomi R, Renn BN, Strong MA, Areán PA. Assessment of real-world use of behavioral health mobile applications by a novel stickiness metric. JAMA Netw Open 2020 Aug 03;3(8):e2011978 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.11978] [Medline: 32744628]
58. Rizvi SL, Dimeff LA, Skutch J, Carroll D, Linehan MM. A pilot study of the DBT coach: an interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. Behav Ther 2011 Dec;42(4):589-600. [doi: 10.1016/j.beth.2011.01.003] [Medline: 22035988]
59. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
60. Schneider RL, Arch JJ, Wolitzky-Taylor KB. The state of personalized treatment for anxiety disorders: A systematic review of treatment moderators. Clin Psychol Rev 2015 Jun;38:39-54. [doi: 10.1016/j.cpr.2015.02.004] [Medline: 25795293]
61. Hedman E, Furmark T, Carlbring P, Ljótsson B, Rück C, Lindefors N, et al. A 5-Year follow-up of internet-based cognitive behavior therapy for social anxiety disorder. J Med Internet Res 2011 Jun 15;13(2):e39 [FREE Full text] [doi: 10.2196/jmir.1776] [Medline: 21676694]
62. Høifødt RS, Mittner M, Lillevoll K, Katla SK, Kolstrup N, Eisemann M, et al. Predictors of response to web-based cognitive behavioral therapy with high-intensity face-to-face therapist guidance for depression: a bayesian analysis. J Med Internet Res 2015 Sep 02;17(9):e197 [FREE Full text] [doi: 10.2196/jmir.4351] [Medline: 26333818]
63. El Alaoui S, Ljótsson B, Hedman E, Kaldo V, Andersson E, Rück C, et al. Predictors of symptomatic change and adherence in internet-based cognitive behaviour therapy for social anxiety disorder in routine psychiatric care. PLoS One 2015 Apr 20;10(4):e0124258 [FREE Full text] [doi: 10.1371/journal.pone.0124258] [Medline: 25893687]
64. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011 Dec;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
65. Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, et al. Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. Cogn Behav Ther 2009 Feb;38(4):206-221. [doi: 10.1080/16506070802408086] [Medline: 19221919]
66. Spek V, Cuijpers P, Nyklícek I, Smits N, Riper H, Keyzer J, et al. One-year follow-up results of a randomized controlled clinical trial on internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years. Psychol Med 2008 May;38(5):635-639. [doi: 10.1017/S0033291707002590] [Medline: 18205965]
67. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: A randomized trial of two versions of CBT for major depression. Behav Res Ther 2010 May;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
68. Craske MG, Niles AN, Burklund LJ, Wolitzky-Taylor KB, Vilardaga JCP, Arch JJ, et al. Randomized controlled trial of cognitive behavioral therapy and acceptance and commitment therapy for social phobia: outcomes and moderators. J Consult Clin Psychol 2014 Dec;82(6):1034-1048 [FREE Full text] [doi: 10.1037/a0037212] [Medline: 24999670]

69. Feske U, Frank E, Kupfer D, Shear M, Weaver E. Anxiety as a predictor of response to interpersonal psychotherapy for recurrent major depression: An exploratory investigation. Depress. Anxiety 1998;8(4):135-141 [FREE Full text] doi: 10.1002/(sici)1520-6394(1998)8:4<135::aid-da1>3.0.co;2-q
70. Young JF, Mufson L, Davies M. Impact of comorbid anxiety in an effectiveness study of interpersonal psychotherapy for depressed adolescents. J Am Acad Child Adolesc Psychiatry 2006 Aug;45(8):904-912. [doi: 10.1097/01.chi.0000222791.23927.5f] [Medline: 16865032]

Abbreviations

AI: artificial intelligence
CFI: comparative fit index
GAD-7: generalized anxiety disorder measure
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial
RMSEA: root mean square of approximation
SRMR: standardized root mean squared residual
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 24.12.20; peer-reviewed by S Schueller, S Amini; comments to author 14.01.21; revised version received 04.02.21; accepted 03.04.21; published 22.06.21.

Please cite as:
Mehta A, Niles AN, Vargas JH, Marafon T, Couto DD, Gross JJ
Acceptability and Effectiveness of Artificial Intelligence Therapy for Anxiety and Depression (Youper): Longitudinal Observational Study
J Med Internet Res 2021;23(6):e26771
URL: https://www.jmir.org/2021/6/e26771
doi:10.2196/26771
PMID:34155984

Original Paper

A Multicomponent mHealth-Based Intervention (SWAP IT) to Decrease the Consumption of Discretionary Foods Packed in School Lunchboxes: Type I Effectiveness–Implementation Hybrid Cluster Randomized Controlled Trial

Rachel Sutherland[1,2,3,4], PhD; Alison Brown[1], BNutrDiet; Nicole Nathan[1,3,4], PhD; Serene Yoong[1,3,4,5], PhD; Lisa Janssen[1], BNutrDiet; Amelia Chooi[1], BNutrDiet; Nayerra Hudson[1], BNutrDiet; John Wiggers[1,3,4], PhD; Nicola Kerr[6], BAppSc; Nicole Evans[7], BPsych; Karen Gillham[1], BSocSci; Christopher Oldmeadow[3], PhD; Andrew Searles[3], PhD; Penny Reeves[1,3], GradDipHlthEcon; Marc Davies[8], BExSci; Kathryn Reilly[1,3,4], PhD; Brad Cohen[9], CA, BCOM-LLB; Luke Wolfenden[1], PhD

[1]Hunter New England Population Health, Wallsend, Australia
[2]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia
[3]Hunter Medical Research Institute, New Lambton Heights, Australia
[4]Priority Research Centre for Health Behaviour, University of Newcastle, Callaghan, Australia
[5]Department of Nursing and Allied Health, School of Health Sciences, Swinburne University of Technology, Hawthorn, Australia
[6]Mid North Coast Local Health District, Port Macquarie, Australia
[7]Central Coast Local Health District, Gosford, Australia
[8]New South Wales Ministry of Health, Liverpool, Australia
[9]SkoolBag, Sydney, Australia

Corresponding Author:
Rachel Sutherland, PhD
Hunter New England Population Health
Booth Building
Longworth Avenue
Wallsend
Australia
Phone: 61 2 4924 6499
Email: rachel.sutherland@health.nsw.gov.au

Abstract

Background: There is significant opportunity to improve the nutritional quality of foods packed in children's school lunchboxes. Interventions that are effective and scalable targeting the school and home environment are therefore warranted.

Objective: This study aimed to assess the effectiveness of a multicomponent, mobile health–based intervention, SWAP IT, in reducing the energy contribution of discretionary (ie, less healthy) foods and drinks packed for children to consume at school.

Methods: A type I effectiveness–implementation hybrid cluster randomized controlled trial was conducted in 32 primary schools located across 3 local health districts in New South Wales, Australia, to compare the effects of a 6-month intervention targeting foods packed in children's lunchboxes with those of a usual care control. Primary schools were eligible if they were not participating in other nutrition studies and used the required school communication app. The Behaviour Change Wheel was used to co-design the multicomponent SWAP IT intervention, which consisted of the following: school lunchbox nutrition guidelines, curriculum lessons, information pushed to parents digitally via an existing school communication app, and additional parent resources to address common barriers to packing healthy lunchboxes. The primary outcome, mean energy (kilojoules) content of discretionary lunchbox foods and drinks packed in lunchboxes, was measured via observation using a validated school food checklist at baseline (May 2019) and at 6-month follow-up (October 2019). Additional secondary outcomes included mean lunchbox energy from discretionary foods consumed, mean total lunchbox energy packed and consumed, mean energy content of core lunchbox foods packed and consumed, and percentage of lunchbox energy from discretionary and core foods, all of which were also measured via observation using a validated school food checklist. Measures of school engagement, consumption of discretionary foods outside of school hours, and lunchbox cost were also collected at baseline and at 6-month follow-up. Data were analyzed via hierarchical linear regression models, with controlling for clustering, socioeconomic status, and remoteness.

Results: A total of 3022 (3022/7212, 41.90%) students consented to participate in the evaluation (mean age 7.8 years; 1487/3022, 49.22% girls). There were significant reductions between the intervention and control groups in the primary trial outcome, mean energy (kilojoules) content of discretionary foods packed in lunchboxes (–117.26 kJ; 95% CI –195.59 to –39.83; *P*=.003). Relative to the control, the intervention also significantly reduced secondary outcomes regarding the mean total lunchbox energy (kilojoules) packed (–88.38 kJ; 95% CI –172.84 to –3.92; *P*=.04) and consumed (–117.17 kJ; 95% CI –233.72 to –0.62; *P*=.05). There was no significant difference between groups in measures of student engagement, consumption of discretionary foods outside of school hours, or cost of foods packed in children's lunchboxes.

Conclusions: The SWAP IT intervention was effective in reducing the energy content of foods packed for and consumed by primary school–aged children at school. Dissemination of the SWAP IT program at a population level has the potential to influence a significant proportion of primary school–aged children, impacting weight status and associated health care costs.

Trial Registration: Australian Clinical Trials Registry ACTRN12618001731280; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=376191&isReview=true

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-019-7725-x

(J Med Internet Res 2021;23(6):e25256) doi:10.2196/25256

KEYWORDS

childhood obesity; lunchboxes; children; child nutrition; mHealth; schools; hybrid; randomized controlled trial; technology

Introduction

Preventing the onset of overweight and obese status in children is a global public health priority [1], given that it impacts negatively on physical health, psychological well-being, and long-term chronic disease risk [2]. The frequent overconsumption of energy-dense, nutrient-poor, or "discretionary" foods throughout childhood, which displace the consumption of core foods consistent with dietary guidelines, is known to be a major contributor to the development of being overweight and obese [3]. Of concern, the poor dietary patterns that are established in childhood track into adulthood and increase the risk of adults being overweight or obese [4]. To address this, the World Health Organization (WHO) recommends implementing populationwide interventions to support the establishment of eating habits in children that are consistent with dietary guidelines [5].

Children consume up to two-thirds of their daily energy intake at school [6]. Consequently, schools have been identified as an optimal setting to implement public health nutrition interventions [5]. Internationally, school-based nutrition research has focused on improving the provision or sale of foods at school canteens [7] or cafeterias [8]. However, in many countries, such as Australia [9], the United Kingdom [10], New Zealand [11], and Denmark [12], a significant proportion of children consume food brought to school from home in a lunchbox. Further, research suggests that the nutritional quality of foods packed in school lunchboxes may be poorer than that available at or provided by schools. For example, in Australia, approximately 5% of items sold at school canteens are discretionary items [13] compared to 40% in children's lunchboxes [14]. A cross-sectional study undertaken in Australia of 1681 students found that lunchboxes contain an average of 3.1 servings of discretionary foods (1200 kJ) and contributed to over 3000 kJ, which is significantly higher than that recommended in dietary guidelines [15]. A further Australian study involving 2143 primary school aged children (mean age 7.96 years) found that just 12% of students' lunchboxes contain only core foods (ie, minimally processed foods recommended in Australian Dietary Guidelines), with a quarter containing 4 or more discretionary servings [16], exceeding the maximum daily amount for children of this age. Similar nutrient compositions have been observed in lunchboxes across the globe, including in New Zealand [11], the United Kingdom [10,17], Canada [18], and the United States [19].

Current evidence regarding the effectiveness of school lunchbox interventions is equivocal. A recent systematic review of such interventions in the school and childcare setting identified just 10 trials and suggested they had little to no effect on the nutritional quality of foods packed or consumed by students [20]. Existing interventions have employed either passive information dissemination strategies to parents, which have limited reach and engagement, or have used intensive face-to-face group-based strategies attracting a biased population group and presenting considerable challenges to implement at scale.

Mobile text messaging– and mobile app–based interventions have been proven to be a scalable and effective approach for improving a variety of health behaviors—including those of parents—to provide a better child diet [21,22]. Our previous pilot study in 12 schools, assessing the feasibility, acceptability, and potential efficacy of the multicomponent SWAP IT intervention [16], used an existing school mobile communication app, along with newly developed school nutrition guidelines, school curriculum, and resources for parents, to encourage a "swap" in their children's lunchboxes of discretionary foods to healthier alternatives consistent with the Australian Dietary Guidelines ("everyday" foods) [23]. The intervention approach was found to be highly feasible to deliver and acceptable to both schools and parents, demonstrating promising short-term improvements in the nutritional quality of foods packed in lunchboxes [16]. Following the encouraging findings of the pilot study, our primary aim was to conduct an adequately powered randomized trial to assess the effectiveness of the SWAP IT multicomponent lunchbox intervention in reducing the kilojoule content from discretionary foods and drinks both packed and consumed by children from school lunchboxes while at school relative to usual care. We also sought to evaluate the effectiveness of the intervention on a range of secondary outcomes, including mean lunchbox energy from discretionary foods consumed, mean total lunchbox energy packed and consumed, mean energy content of core lunchbox foods packed and consumed, percentage of lunchbox energy from discretionary and core foods, measures of school engagement, consumption of discretionary foods outside of school hours, and lunchbox cost.

Methods

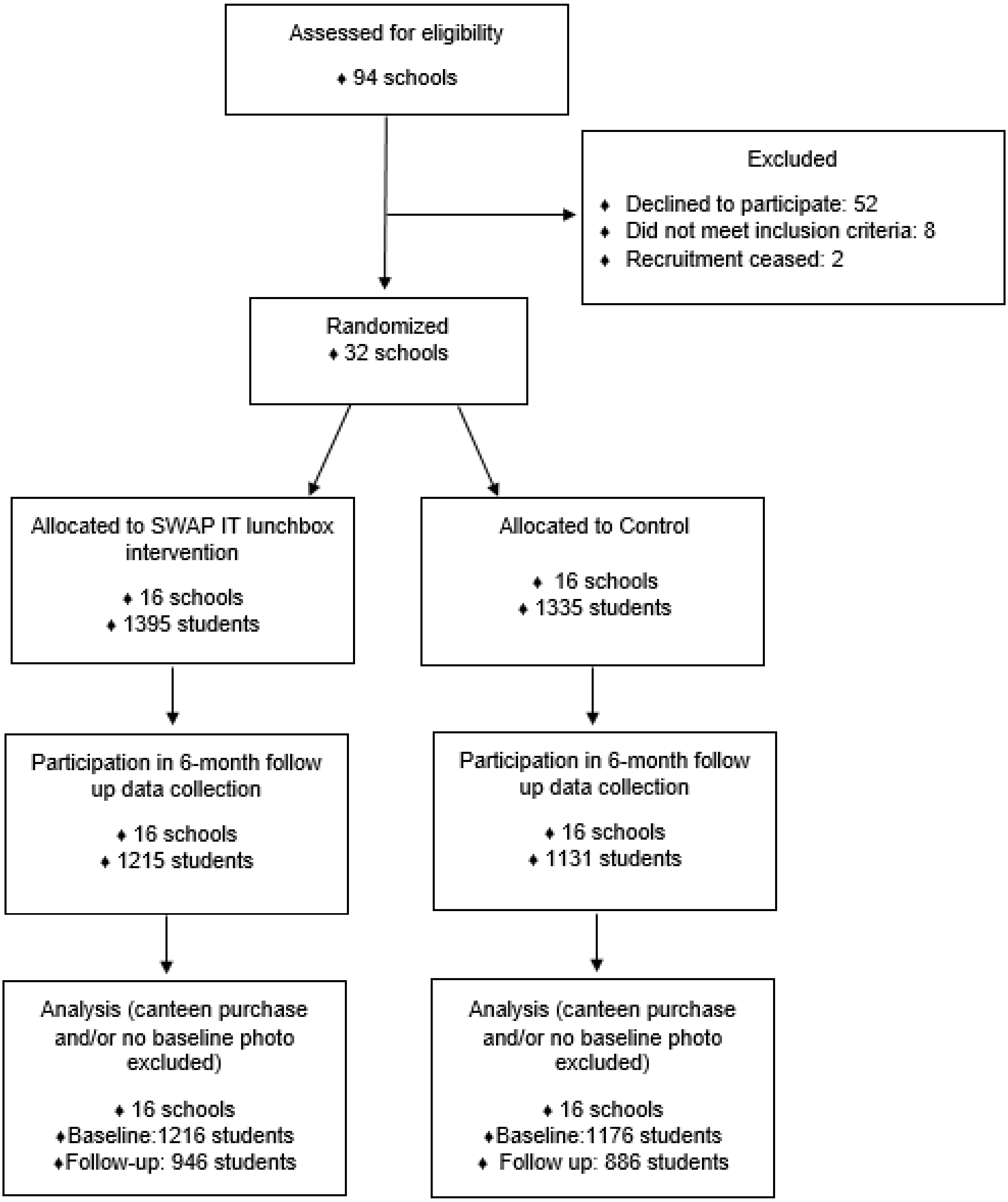
Ethics and Registration

The research was conducted and reported in accordance with the requirements of the Consolidated Standards of Reporting Trials (CONSORT) statement [24]. Approval to conduct this study was obtained from the Hunter New England Human Research Ethics Committee (reference #06/07/26/4.04), University of Newcastle (reference #H-2008-0343) and the New South Wales (NSW) State Education Research Applications Process (#2018247) and was prospectively registered with Australian New Zealand Clinical Trials Register (#12618001731280). A detailed description of the methods and intervention are outlined in the study protocol [25].

Study Design and Setting

A type I effectiveness–implementation hybrid cluster randomized controlled trial was conducted with 32 primary (students aged approximately 5-12 years) schools across 3 local health districts in NSW, Australia (Figure 1). Schools were randomized to receive a 6-month (2 school terms), multicomponent lunchbox intervention or a usual care control arm (16 schools per arm). Outcome assessments were conducted in a cohort of students at baseline and at 6 months after randomization. The primary outcome was mean energy (kilojoules) content of discretionary lunchbox foods and drinks packed in lunchboxes assessed via lunchbox observation. Other registered outcomes related to implementation processes, including intervention acceptability, appropriateness and feasibility [26,27], cost-effectiveness of the intervention, and impact on mean daily nutrient consumption, will be reported separately.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) flow diagram.



Sample and Participants

Schools

Schools were considered eligible if they met the following criteria: government primary schools catering for students from kindergarten to year 6 and located in one of the participating local health districts, greater than 120 student enrolments, current users of the preferred school mobile communication app (SkoolBag), and not participating in other nutrition-based research studies. Schools purchase the communication app for a nominal fee annually, which is then free for parents to download, enabling direct school–parent communication. The app is used by approximately 60% of schools in the region. Central schools (catering for students aged 5-18 years) and schools primarily catering for children with additional needs (such as intellectual disabilities) were excluded. According to a random number generator in Excel (Microsoft Corporation), eligible schools meeting the above criteria were sent a letter of invitation in random order. One week following the invitation, a member of the research team contacted the school principal via telephone to seek consent. A face-to-face meeting was offered to all schools to outline the requirements of the study. Recruitment and consent of schools occurred between February 2019 and May 2019. Recruitment continued until 32 schools provided active signed principal consent to participate.

Parents and Students

Opt-in parental consent was required for children and parents to participate in the evaluation of the behavioral outcomes. Parents were also required to be active users of the school communication app, defined as downloading the school communication app on the parent consent form. A strategy to recruit parents and students was developed based on the pilot study and reviews of evidence for facilitating participation in school-based research [16,28]. Following principal consent, all parents with a child enrolled in classes from kindergarten to year 6 (5-12 years) were invited to participate in the study evaluation measures, which included a lunchbox observational assessment, parent survey, and student survey (year 5 and 6 students). Students were provided with an information package outlining the study and a consent form. Parents were asked via the consent form if they were an active user (ie, downloaded the app) of the school communication app. One week after the information package was distributed, parents who had not returned a consent form were telephoned by school-employed staff. A replacement consent form was distributed via mail to parents who provided verbal consent over the phone.

Randomization and Blinding

Following baseline data collection, schools (cluster) were randomly allocated in a 1:1 ratio to the intervention or control group based on a random number function in Excel. Randomization was undertaken by a statistician not involved in contacting schools in the study intervention or assessment and stratified by the socioeconomic status of school locality using the Socio-Economic Indexes for Areas (SEIFA 2016), as socioeconomic status is associated with lunchbox contents and child diet [29,30]. Research personnel involved in data collection and lunchbox content analysis were blind to group allocation, as all identifiable school information was removed prior to data analysis. Data collection staff were not informed of group allocation; however, this might have been disclosed to them by school staff during field activity. Due to the inability to conceal intervention delivery, school personnel were notified of their group allocation via a phone call.

Multicomponent Intervention

The multicomponent intervention based on the previous pilot was codeveloped by a multidisciplinary team comprising academic and end-user stakeholders from government health agencies, educational systems, universities, and technology partners and included parent representatives with expertise in nutrition, school-based health interventions, behavior change, implementation science, and technology-based interventions.

Conceptual Framework

The Behaviour Change Wheel [31] was used to guide the development of the intervention. Extensive formative research encompassing a review of published literature; focus groups with parents to identify local contextual barriers; telephone interviews with parents (n=228) (L Janssen, R Sutherland, and N Nathan; unpublished data, 2019) and principals (n=196) [32] to assess barriers, acceptability of intervention strategies, and content and delivery mode; and a literature review of existing lunchbox interventions [20] were undertaken to select behavior change techniques and strategies to support parents to pack healthy school lunchboxes. Multimedia Appendix 1 outlines the Behaviour Change Wheel mapping process and outlines the chosen behavior change techniques incorporated into the SWAP IT intervention.

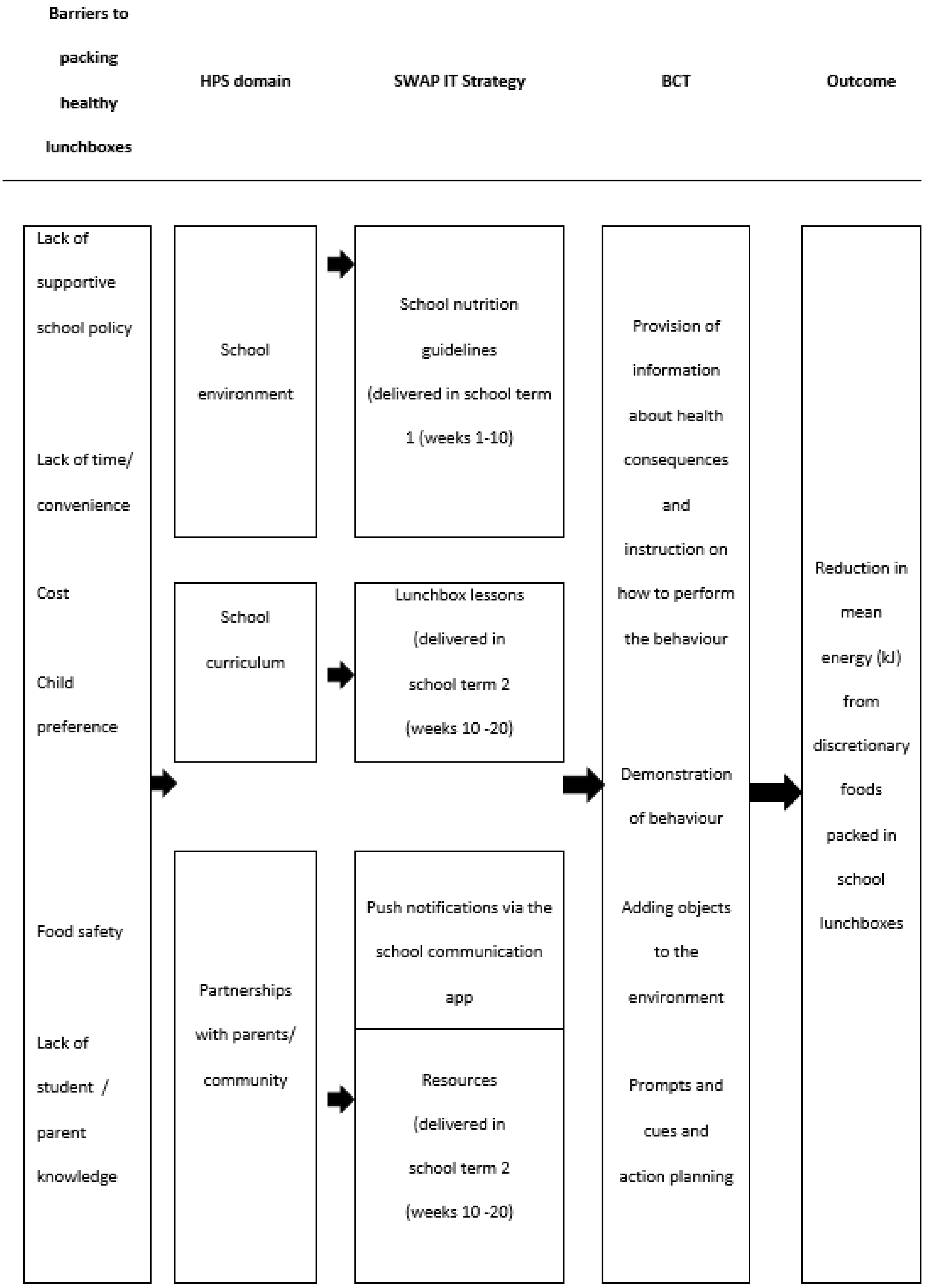
Figure 2 provides an overview of the SWAP IT intervention logic. The SWAP IT intervention encouraged lunchbox “swaps” from discretionary food items to Australian Dietary Guideline–based healthier alternatives known as “everyday” foods. The multicomponent lunchbox intervention consisted of 4 strategies outlined in the following section. The mobile health component included weekly pushed messages to parents delivered via an existing school mobile communication app, SkoolBag, in addition to embedding lunchbox content within the app for parents to access. A detailed 4-part description of the intervention has been published in a protocol [25], and the intervention included the 4 following strategies:

1. Lunchbox nutrition guidelines: Using a template developed by the project team, school principals developed, endorsed, and disseminated nutrition guidelines to parents which were consistent with the WHO and the NSW Department of Education Nutrition in Schools policy [33]. Guidelines were disseminated to parents in the first 5 weeks of the intervention via the SkoolBag app and school newsletters to demonstrate schools’ endorsement of the SWAP-IT program.
2. Weekly pushed lunchbox messages: Through the SkoolBag app, 10 weekly electronic messages (push notifications) to support the packing of healthy lunchboxes were disseminated to parents or caregivers. Messages were codeveloped by the research team, public health nutritionists, health promotion practitioners, teachers, and parents and were optimized and refined via a study involving 511 parents [34]. The distribution of the messages via the school communication app was managed centrally by the project team. This allowed all parents at the schools allocated to the intervention group who had downloaded the app to receive the pushed messages via the research team and prevented the need to rely on each individual school to push the weekly content to parents. This centrally coordinated effort therefore did not require school time or resources and thereby maximized the fidelity of the intervention. The pushed messages aligned to parent-reported barriers to packing healthy school lunchboxes: lack of time or convenience, knowledge of suitable swaps, child preference, cost, food safety, and lack of school nutrition policy. Where possible, a swap within the same food category was suggested (eg, for packaged foods). The pushed messages were designed to act as prompts and cues to reinforce packing of everyday foods. The messages were connected to embedded videos

developed by the research team to align to parent-reported barriers which provided tips and suggestions to assist parents to pack everyday foods that were quick, convenient, and low cost and to connect parents with tools and resources to improve their knowledge and skills to swap out discretionary foods and pack everyday foods.

3. Resources for parents: Links embedded in the app messages connected parents with electronic resources housed on the program website. These resources provided information regarding health consequences, simple healthy lunchbox swaps that addressed child preference, cost, convenience, and food safety. Physical resources, including a SWAP IT ideas booklet (lunchbox ideas), clear drink bottle for water, and an ice brick to support food safety, were also provided to parents and were distributed to students and parents via the schools' usual methods of dissemination.
4. Curriculum resources for schools: Schools were provided with a short online teacher professional learning module (10 minutes) developed by the research team, which included public health nutritionists, health promotion practitioners, and teachers outlining the rationale for the study and providing the skills and resources required to deliver the classroom curriculum lessons. Schools were also provided stage-appropriate curriculum resources which were codeveloped by the research team with input from teachers, parents, and education partners to align with syllabus outcomes that were developed by dietitians and teachers in order to reinforce healthy food preferences. This required teachers to deliver 3 curriculum lessons 10 minutes in duration. Curriculum resources were designed to address the identified barrier to packing a healthy lunchbox of "child preference for discretionary foods."

Figure 2. SWAP IT logic model. BCT: behaviour change technique; HPS: health promoting schools framework.



Control Schools

Schools allocated to the control group had access to the SkoolBag app but not the lunchbox intervention content. The SWAP IT website was freely accessible by the general public, including parents and schools; however, schools and parents were not notified or directed to this site. There was no information (nutrition or otherwise) provided to the control group, and they participated in data collection only and continued usual school business.

Data Collection and Measures

Lunchbox Energy

The primary outcome was the mean energy (kilojoules) content of discretionary foods packed in the school lunchboxes by parents who were users of the school mobile app, assessed at baseline and at 6-month follow-up. A detailed description of the study measures and data collection methods have been described in a published protocol [25]. Lunchbox energy content was assessed from photos of lunchboxes taken at school by trained research assistants prior to the first meal break with a valid and reliable lunchbox observational audit, known as the School Food Checklist (SFC) [35,36]. The SFC is a previously validated tool shown to be accurate and reliable in measuring energy from food and drinks for the Australian context. The SFC [35,36] enabled the assessment of the kilojoule content and serving size for each lunchbox item. Two trained dietitians observed each school lunchbox photo and classified each food and drink item according to its SFC category as “everyday foods” or “discretionary foods” and assessed the kilojoule content and serving size for each lunchbox item and the serving size. The checklist included 20 food and drink categories, including main food items, such as bread, fast food, and leftovers/mixed dishes; and snack items such as noodles, packaged snacks, biscuits and crackers, chocolate and candy, cheese, eggs, dried fruit and nuts, muesli and fruit bars, cakes and buns, muffins and scones, pastries, desserts, yoghurt, fruit, vegetables, milk, soft drink, and water and fruit juice. “Everyday” items referred to food and drink items that were part of the core food groups as determined by the Australian Dietary Guidelines [23]. Food items classified as “discretionary” were items considered to be energy dense with minimal nutritional value, including cakes, chocolate, candy, chips, muesli bars, and fast food [23]. The serving size of each lunchbox item and kilojoules per serving information was obtained from FoodWorks Professional Edition V7 (version 7, Xyris Software). To further aid this process, decision rules developed in the previous study [16] were used to ensure standardization of assessments.

The secondary outcomes associated with lunchbox energy were mean total energy (kilojoule) packed within the lunchbox; mean total energy (kilojoules) consumed from the lunchbox; mean energy (kilojoules) from discretionary foods and drinks consumed within the lunchbox; mean energy (kilojoules) from healthy foods packed and consumed from the lunchbox; and percentage of lunchbox energy from discretionary and healthy foods and drinks, both packed and consumed. Data were collected at baseline and immediately after the 6-month intervention with the SFC as outlined in the previous section. Following the analysis of the premeal lunchbox photo, dietitians analyzed the postmeal photo.

Student Consumption of Discretionary Foods Outside of School Hours

At baseline and at follow-up, parents were asked to report, via a short telephone survey, on their child’s intake of discretionary foods outside of school hours and on weekends to identify any compensatory nutrition behavior occurring outside of school hours. Measures were taken from the NSW Schools Physical Activity and Nutrition Survey [37]. Parents reported on the following 6 categories of discretionary foods: (1) fried potato products, (2) potato chips and other salty snacks, (3) sweet biscuits and cakes, (4) confectionary, (5) ice cream or ice blocks, and (6) fruit juice. The frequency of consumption for consenting students was reported at baseline and immediately after the intervention at 6 months.

Procedures

To assess the foods packed in the lunchbox (premeal assessment) on a randomly selected school day prior to recess, at lunch, or during in-class vegetable and fruit breaks [38,39], consenting students were asked to display the contents of their lunchbox on their desk in the classroom. Parents and students were not informed of the exact day of data collection. A preprepared paper grid was placed under the lunchbox contents and used to assess the scale and serving size of the items. Any foods not easily identified were discussed with the student and further details were recorded on the grid paper prior to being photographed. The photo was taken by trained research assistants prior to any foods being consumed. Students were asked if they intended to purchase food from the canteen that day, and if so, were removed from the analysis.

To assess the consumption of foods packed in the lunchbox (postmeal assessment), on the same day, students were asked to keep all unconsumed or partially consumed food items in their lunchboxes. Following all meal breaks, students were asked to place unconsumed or partially consumed items from their lunchbox onto the grid paper, and a second photograph of all remaining food was taken. Measures relating to consumption were based on the second photograph of the day being taken after all meal breaks had occurred and all uneaten food had been placed back into the lunchbox container. Consumption was calculated by subtracting the postmeal assessment from the premeal assessment.

Trained dietitians, blinded to group allocation, observed each school lunchbox photo in order to classify each food and drink item according to its SFC category and the serving size. All lunchbox photos were assessed by 2 dietitians working together to make a consensus decision on the analysis for each lunchbox. To further aid this process, decision rules were developed to ensure standardization of assessments. Differences in opinion between dietitians were resolved following consultation with a third dietitian assessor. Following the analysis of the premeal lunchbox photo, dietitians then analyzed the postmeal photo. Energy consumption was calculated by subtracting the energy content of foods and drinks remaining in students’ lunchboxes at the postmeal assessment from the energy content of foods and drinks in the lunchbox during premeal assessment (“foods consumed”).

Student School Engagement

We also assessed impact on engagement, as research suggests that improved nutrition correlates with greater school attendance, improved concentration, and higher academic achievement [40]. At baseline and at follow-up, students in years 5 and 6 completed selected items from the validated School Engagement Measure via a pen and paper survey. The School Engagement Measure is a 19-item survey that provides a measure of students’ behavioural (5 items), emotional (6 items), and cognitive engagement (8 items) at school, which are outcomes considered important for achieving positive academic outcomes [41].

Student Consumption of Discretionary Foods Outside of School Hours

To ensure any reduction in energy intake occurring while at school did not result in compensatory intake outside of school hours (potential adverse event), parents were asked via a short telephone survey at baseline and at follow-up to report on their eldest eligible child's intake of discretionary foods outside of school hours and on weekends. Measures were taken from the NSW Schools Physical Activity and Nutrition Survey [37]. Parents reported on 6 categories of discretionary foods, including (1) fried potato products, (2) potato chips and other salty snacks, (3) sweet biscuits and cakes, (4) confectionary, (5) ice cream or ice blocks, and (6) fruit juice and reported the frequency of consumption as never or rarely, 1 to 2 times per week, 3 to 4 times per week, 5 to 6 times per week, once per day, or 2 or more times per day.

Lunchbox Cost

It has been hypothesized that one potential adverse effect of encouraging healthier lunchbox swaps is increased family financial burden due to the potential higher cost of healthier products [42]. To assess this, the mean cost of lunchbox items before and after intervention was assessed via the SFC and were compared between intervention and control groups at baseline and at follow-up to determine if the intervention resulted in any adverse financial effects for families. Costing was determined with an average of prices from foods within the category accessed from a local retail audit of similar foods as of October 2018.

Statistical Analysis

Analyses were conducted using SAS version 9.3 (SAS Institute) from January 2020 to June 2020. School and student characteristics were summarized for intervention and control schools. Summary statistics are used to describe all variables of interest. Students that resided in postcodes ranked in the top 50% of state postcodes based on the 2016 SEIFA [30] were categorized into "higher socioeconomic areas," whereas those in the lower 50% were categorized into "lower socioeconomic areas." Students' postcodes were also used to categorize their locality as either "rural" (those schools in outer regional, remote, or very remote areas) or "urban" (those in regional or major cities) based upon the 2016 Accessibility/Remoteness Index of Australia [43].

The differences between groups in the primary and secondary outcomes were assessed using hierarchical linear (or logistic for binary outcomes) regression models. Models were adjusted for SEIFA, remoteness, and baseline values, and a random level intercept for schools was included to adjust for the clustered design of the study. Analysis followed intention-to-treat principles, where schools and students were analyzed according to their randomized treatment allocation. All statistical tests were 2-tailed with an α of .05. As specified in the study protocol [25], data were analyzed only for students whose parents had reported downloading the required SkoolBag app to ensure exposure to the intervention, and students intending to purchase food or drinks from the canteen or who did not bring lunch were removed from the primary analysis to focus on students whose lunchbox was their source of energy for the day [16].

Sample Size and Power

According to our pilot results [16], a standard lunchbox contains 1089 kJ (SD 900 kJ) of discretionary foods. With an intraclass correlation coefficient of 0.05, 32 schools with 140 students per school enabled detection of a 200-kJ difference between groups at follow-up on the primary trial outcome, with 80% power at a significance level of $P<.05$. As approximately 420 kJ across a whole day has the potential to reduce the prevalence of childhood obesity [44,45] and as it is recommended that a child consumes a third of their daily energy requirements while at school [9], this magnitude of effect was considered meaningful at a population level.

Results

Sample

A sample of 94 schools was assessed for eligibility to participate in the study, and 91 were approached in order to obtain the quota of 32 consenting schools (35.2%). Consenting and nonconsenting schools were similar in geographic location, size, and school socioeconomic status, with the 32 consenting schools enrolling a total of 7212 students (or 5048 families). Of these, 3022 provided parental consent to participate in the lunchbox observation to evaluate the outcomes of the study (41.90%). From the 3022 consenting students, 2730 (1395 intervention and 1335 control) lunchboxes were observed at baseline and 2346 (1215 intervention, 1131 control) at follow-up, with the discrepancy being due to student absences and school events or excursions. Table 1 outlines the school and student characteristics of those consenting to participate. At baseline, the consenting schools and students allocated to the intervention and control groups had similar characteristics; however, the intervention group had a higher proportion of schools located in disadvantaged areas.

Table 1. Sample characteristics of schools and students at baseline.

Characteristics	Intervention	Control
Schools		
Allocation, n	16	16
Location, n (%)		
Urban	8 (50.0)	6 (37.5)
Rural	8 (50.0)	10 (62.5)
Socioeconomic status[a], n (%)		
Most disadvantaged	13 (81.2)	13 (81.2)
Least disadvantaged	3 (18.8)	3 (18.8)
Schools with greater than 10% Aboriginal or Torres Strait Islander student enrolments, n	10	11
Students		
Allocation, n	1216	1176
Sex, n (%)[b]		
Female	592 (50.04)	550 (48.37)
Male	591 (49.96)	587 (51.63)
Mean age (years)	7.88	7.68
Socioeconomic status[a], n (%)		
Most disadvantaged	938 (77.14)	789 (67.09)
Least disadvantaged	278 (22.86)	387 (32.91)

[a] Socioeconomic status is based on SEIFA Index of relative socioeconomic disadvantage 2016: most disadvantaged = lowest quartiles of SEIFA; least disadvantaged = highest quartiles of SEIFA.

[b] Information on sex missing for 72 students.

Primary Outcome: Mean Energy (Kilojoules) Content of Discretionary Foods Packed From the School Lunchboxes

At 6-month follow-up, the difference between the intervention and control group in the mean energy (kilojoules) content of discretionary foods packed in school lunchboxes was –117.71 kJ (95% CI –195.59 to –39.83; *P*=.003). A sensitivity analysis on the primary outcome using complete cases indicated a similar result of –120.43kJ (95% CI –200.82 to –40.04; *P*=.005).

Secondary Lunchbox Energy Outcomes

The mean total energy (kilojoules) packed in lunchboxes (–88.38 kJ; 95% CI –172.84 to –3.92; *P*=.04) and mean total energy (kilojoules) consumed from lunchboxes (–117.17kJ; 95% CI –233.72 to –0.62; *P*=.05) both reduced in favor of the intervention group. There was also a significant reduction in percentage of lunchbox energy packed from discretionary foods between groups (–3.16%; 95% CI –5.46 to –0.86; *P*=.01), while the percentage of lunchbox energy from everyday foods increased (3.16%; 95% CI 0.86-5.46; *P*=.01). A significant reduction favoring the intervention group in the mean energy (kilojoules) from discretionary foods consumed from lunchboxes (–96.31kJ; 95% CI –194.63 to 2.01; *P*=.05) was also observed. There was no statistical difference between groups in the mean lunchbox energy from everyday foods (kilojoules) packed in lunchboxes (32.85 kJ; 95% CI –31.61 to 97.31; *P*=.31) or consumed (–21.91 kJ; 95% CI –112.38 to 68.56; *P*=.62). Table 2 outlines the lunchbox energy packed and consumed by group. Multimedia Appendix 2 outlines the food and drink items packed in lunchboxes.

Table 2. Mean energy and percentage of energy from everyday and discretionary foods packed and consumed from student lunchboxes.

Outcome	Intervention		Control		Difference in energy between groups at follow-up, mean (95% CI)	*P* value
	Baseline mean, (SD) (n=1216)	Follow-up, mean (SD) (n=946)	Baseline, mean (SD) (n=1176)	Follow-up, mean (SD) (n=886)		
Daily energy (kilojoules) packed in student lunchboxes						
Primary outcome: lunchbox energy from discretionary foods packed in lunchboxes	1214.86 (876.49)	1156.77 (841.76)	1067.38 (898.82)	1105.06 (859.06)	–117.26 (–195.59 to 39.83)	.003
Lunchbox energy from everyday foods packed in lunchboxes	1616.19 (628.34)	1610.93 (624.41)	1644.17 (621.73)	1605.81 (610.02)	32.85 (–31.61 to 97.31)	.31
Total lunchbox energy packed in lunchboxes	2831.05 (927.81)	2767.70 (873.52)	2711.54 (962.33)	2710.87 (878.44)	–88.38 (–172.84 to –3.92)	.04
Daily lunchbox energy (kilojoules) consumed by students						
Lunchbox energy from discretionary foods consumed from lunchboxes	901.30 (745.60)	876.70 (717.23)	744.19 (717.20)	802.75 (677.23)	–96.31 (–194.63 to 2.01)	.05
Lunchbox energy from everyday foods consumed from lunchboxes	1270.85 (631.79)	1282.56 (622.95)	1304.69 (600.58)	1341.72 (607.53)	–21.91 (–112.38 to 68.56)	.62
Total lunchbox energy consumed from lunchboxes	2172.15 (895.82)	2159.26 (810.78)	2048.88 (853.84)	2144.48 (743.22)	–117.17 (–233.72 to –0.62)	.05
Lunchbox energy coming from discretionary and everyday foods (%)						
Packed lunchbox energy from discretionary foods	40.10 (23.31)	39.04 (23.94)	35.84 (23.69)	37.90 (23.81)	–3.16 (–5.46 to –0.86)	.01
Packed lunchbox energy from everyday foods	59.90 (23.31)	60.96 (23.94)	64.16 (23.69)	62.10 (23.81)	3.16 (0.86 to 5.46)	.01
Total cost (Aus $) of lunchbox items	3.94 (1.35)	3.91 (1.36)	3.78 (1.38)	3.78 (1.32)	–0.06 (–0.18 to 0.07)	.37

Student Engagement

Table 3 outlines the student engagement measures. There were no observed differences between groups for any measure of student engagement after the 6-month intervention, including for student total school engagement measure score (–0.08; 95% CI –0.18 to 0.02; *P*=.10), student behaviour (–0.05; 95% CI –0.15 to 0.04; *P*=.24), or emotional (–0.08; 95% CI –0.2 to 0.06; *P*=.26) or cognitive engagement (0.09; 95% CI –0.22 to 0.05; *P*=.20).

Table 3. Mean school engagement measure by group at baseline and at follow-up.

Mean school engagement score	Intervention		Control		Difference in engagement between groups at follow-up, mean (95% CI)	*P* value
	Baseline, mean (SD) (n=364)	Follow-up, mean (SD) (n=309)	Baseline, mean (SD) (n=299)	Follow-up, mean (SD) (n=241)		
Behavior score	4.12 (0.59)	4.09 (0.62)	4.11 (0.65)	4.14 (0.66)	–0.05 (–0.15 to 0.04)	.24
Emotion score	3.55 (0.91)	3.33 (0.99)	3.56 (0.92)	3.40 (0.98)	–0.08 (–0.22 to 0.06)	.26
Cognitive score	2.92 (0.87)	2.80 (0.87)	2.87 (0.83)	2.83 (0.85)	–0.09 (–0.22 to 0.05)	.20
Total school engagement	3.44 (0.66)	3.31 (0.71)	3.42 (0.68)	3.35 (0.70)	–0.08 (–0.18 to 0.02)	.10

Student Consumption of Discretionary Foods Outside of School Hours

There were no differences between groups in the foods consumed outside of school hours, indicating no compensatory consumption of discretionary foods outside of care.

Lunchbox Cost

The total cost of lunchbox foods following the intervention did not differ between groups (–Aus $0.06; 95% CI –0.18 to 0.07; *P*=.37; Table 2).

Discussion

This trial investigated the effectiveness of the SWAP IT intervention on the energy of students' lunchbox foods, both packed and consumed, using an existing school communication app provided directly to parents. Relative to lunchboxes in the control group, the lunchboxes in the intervention group contained significantly less mean energy from discretionary foods corresponding to 117 kJ per day or a 600 kJ reduction over a school week. The SWAP IT intervention also resulted in a reduction in mean energy from discretionary foods that were consumed by students (96.31 kJ). The mean total lunchbox energy both packed and consumed was also significantly less in intervention lunchboxes, and the percentage of energy from discretionary foods decreased by 3.16%, while percentage energy from everyday foods correspondingly increased. The lunchbox energy coming from everyday foods that were consistent with dietary guidelines did not statistically differ between groups, indicating the change in total energy observed was primarily from a reduction in discretionary foods. These favorable nutrition outcomes occurred while the cost of packing a lunchbox remained stable across groups, indicating the changes made to lunchboxes did not result in additional costs. The intervention, however, did not result in changes to student school engagement at school.

Although it is challenging to make direct comparisons, the magnitude of reduction in energy from discretionary foods appears favorable compared to previous lunchbox interventions. Of the 10 included studies within a systematic review of lunchbox interventions conducted within the school and childcare environment [20], 4 targeted the packing of discretionary foods, with evidence for the effectiveness of interventions on what is packed in lunchboxes in relation to discretionary foods, sugar-sweetened drinks, or other core foods being equivocal. Of the 2 studies conducted in the school environment, results were mixed, with interventions impacting on the reduction of either high fat salty snacks (reduction of 2.8 gm of savory snacks; *P*=.04) or sweet confectionary, fruit drinks, or candy (–0.43 servings; *P*=.001), but not both [20]. Furthermore, only 1 study within this review used a similar methodology of assessment based on lunchbox photography and observation to estimate serving size [20]. The study found no significant effect on the nutritional quality of food brought from home, which might have been due to a lack of power or a result of the complex information dissemination pathway of their intervention that involved sending messages to parents via newsletters and lessons delivered to children. To target behavior change, our research team conducted extensive formative assessment including mapping barriers and consulting with key stakeholders to co-design the SWAP IT using a theoretical framework. The multicomponent program also dually targeted both the school, via guidelines and curriculum, and parents, via direct messages, which may explain the favorable intervention effect.

To improve health at a population level, interventions shown to be effective under research conditions need to be scaled up to reach a large proportion of the population [46]. Few school-based behavioral interventions have been suggested to be suitable for large-scale dissemination, as they require expertise and resources not readily available within schools and often use high-intensity delivery modes, such as face-to-face training [20]. The use of digital delivery modes overcome many of these barriers; however, poor adoption and ongoing engagement with new apps or websites, for example, are often an impediment to population-level reach and improved health gains [47]. To address this, the SWAP IT intervention adopted a multicomponent intervention design that addressed many of the existing limitations [20]. Incorporating the SWAP IT behavioral intervention components as a complement into an existing school communication app that had already been adopted by schools and downloaded by parents was undertaken to overcome the challenges of both population-level reach and digital engagement. A similar approach, in which a nutrition intervention was embedded into an existing online school canteen ordering system, also resulted in a significant intervention effect on energy, sugar, and fat [48]. This suggests that embedding digital interventions within existing systems supported by additional behavior change strategies may be superior to developing and implementing new digital health interventions alone.

Although a reduction in energy from discretionary foods of 600 kJ per week may appear small at an individual level, at a population level, it has the potential to lower the risk of individuals being overweight or obese, result in a gain of health-adjusted life years, and make a significant contribution toward savings in health care costs [49]. Given its potential reach, with 86% of students taking a packed lunch on a daily basis [9] and 90% of those students packing at least 1 serving of discretionary food in their lunchbox [14], this intervention has the potential to immensely shift the consumption of discretionary foods. Further, as 60% to 70% [50] of schools in NSW Australia and the United States, respectively, already using school communication apps, interventions such as SWAP IT have the potential to reach millions of parents on a daily basis. Further investigation evaluating the cost-effectiveness and implementation process of the SWAP IT intervention is needed to confirm if the SWAP IT intervention warrants large scale dissemination. Future research should focus on developing strategies that maximize the adoption or uptake of the SWAP IT intervention by schools at scale and methods for sustaining school engagement to continue the impact on parent behavior change.

The results of this trial should be interpreted within the context of its strengths and limitations. Study strengths include the experimental hybrid design, with randomized controlled trials being considered the gold standard for evaluating causal effects of interventions. The SWAP IT trial was also developed using behavior change theory and used direct observation and validated tools to assess lunchbox contents, which strengthened the ability of the study to accurately measure the true impact of the study outcomes. Although the effect size of the SWAP IT effectiveness trial was smaller than that of the previous pilot [16], the significant results were replicated, indicating that, pending further evaluation exploring the implementation outcomes and cost effectiveness, the intervention warrants consideration for large-scale dissemination. However, a number of limitations should be considered. The trial had a lower than anticipated participation and consent rate from schools and particularly parents, with only 41.90% of parents consenting to participate in the lunchbox observations. Upon enquiry, we believe this is primarily due to the measurement component, in which lunchbox observations might have been considered an encroachment on privacy [51], given that the acceptability of the intervention for schools and parents was high at 84% [16]. The intervention was also multicomponent, and isolating the impact of each individual strategy was not possible in this trial. Although the large-scale dissemination of pushed messages to parents via the app is feasible and has high fidelity, implementation of school-level strategies may require additional support. This trial had a follow-up period of 6 months, and the long-term sustainability of the intervention in both schools and with parents is unknown. Further investigation is warranted to ensure the intervention has an ongoing desirable impact on lunchbox behavior.

The SWAP IT intervention presents an effective digital behavior change solution to a large and long-standing public health problem of a high consumption of discretionary foods by children while at school. Given the significant impact on lunchbox food energy that has been demonstrated by the previous pilot trial and replicated in this effectiveness trial at a larger scale, the intervention provides an attractive option to policy makers to complement existing public health programs targeting the school nutrition environment. Following further evaluation to determine its implementation process, outcomes, and cost-effectiveness, models to further scale up and maximize the adoption of SWAP IT will ensure that a public health benefit can be realized.

Acknowledgments

The authors wish to acknowledge the research assistants involved and the schools participating in this study, as well as the Department of Education, Hunter Region, for their contribution to the advisory group.

This project was funded by the NSW Ministry of Health, Translational Research Grant Scheme. The NSW Ministry of Health has not had any role in the design, data collection, analysis of data, interpretation of data, or dissemination of the study. The project also received infrastructure support from the Hunter Medical Research Institute. RS is supported by a National Health and Medical Research Council Translating Research into Practice Fellowship (#APP1150661) and a Hunter New England Clinical Research Fellowship; LW is supported by a National Health and Medical Research Council Career Development Fellowship (#APP1128348) and a Heart Foundation Future Leader Fellowship; SY is supported by a Discovery Early Career Researcher Award fellowship.

Authors' Contributions

RS and AB led the development of this manuscript. RS, NN, SY, LJ, and LW conceived the intervention concept. RS and LW secured funding for the study. RS, NN, and LW guided the design and piloting of the intervention. RS, AB, LW, NN, LJ, and JW guided the evaluation design and data collection. CO developed the analysis plan. RS, AB, NN, LJ, JW, NK, NE, CO, AS, PR, CR, BS, MD, KR, BC, KG, and LW are all members of the advisory group that oversee the program and monitor data. RR, AW, NH, AB, LJ, and AC are all members of the project team that oversee the implementation and evaluation of the program. All authors contributed to developing the protocols and reviewing, editing, and approving the final version of the paper.

Conflicts of Interest

Authors RS, NN, LW, KG, NE, and JW receive salary support from their respective local health districts. Hunter New England Local Health District contributes funding to the project outlined in this protocol. None of these agencies were involved in the peer review of this grant. RS and NN are associate editors for BMC Public Health. All other authors declare that they have no competing interests.

Multimedia Appendix 1
Using the Behaviour Change Wheel process to map barriers to packing healthy lunchboxes with identified intervention functions and suitable behaviour change techniques (BCTs).
[DOCX File , 16 KB - jmir_v23i6e25256_app1.docx]

Multimedia Appendix 2
Food and drink items packed in lunchboxes.
[DOCX File , 15 KB - jmir_v23i6e25256_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1164 KB - jmir_v23i6e25256_app3.pdf]

References

1. Report of the Commission on Ending Childhood Obesity: implementation plan: executive summary. World Health Organization. 2017. URL: https://www.who.int/end-childhood-obesity/publications/echo-plan-executive-summary/en/ [accessed 2021-05-01]
2. Lobstein T, Baur L, Uauy R. Obesity in children and young people: a crisis in public health. Obes Rev 2004 May;5 Suppl 1:4-104. [doi: 10.1111/j.1467-789X.2004.00133.x] [Medline: 15096099]
3. Chung A, Peeters A, Gearon E, Backholer K. Contribution of discretionary food and drink consumption to socio-economic inequalities in children's weight: prospective study of Australian children. Int J Epidemiol 2018 Jun 01;47(3):820-828. [doi: 10.1093/ije/dyy020] [Medline: 29514246]
4. Venn AJ, Thomson RJ, Schmidt MD, Cleland VJ, Curry BA, Gennat HC, et al. Overweight and obesity from childhood to adulthood: a follow-up of participants in the 1985 Australian Schools Health and Fitness Survey. Med J Aust 2007 May 07;186(9):458-460. [doi: 10.5694/j.1326-5377.2007.tb00997.x] [Medline: 17484707]
5. School policy framework: implementation of the WHO global strategy on diet, physical activity and health. World Health Organization. 2008. URL: https://apps.who.int/iris/handle/10665/43923 [accessed 2021-05-01]
6. Sharma M. School-based interventions for childhood and adolescent obesity. Obes Rev 2006 Aug;7(3):261-269. [doi: 10.1111/j.1467-789X.2006.00227.x] [Medline: 16866974]
7. Wolfenden L, Nathan N, Williams CM, Delaney T, Reilly KL, Freund M, et al. A randomised controlled trial of an intervention to increase the implementation of a healthy canteen policy in Australian primary schools: study protocol. Implement Sci 2014 Oct 11;9:147. [doi: 10.1186/s13012-014-0147-3] [Medline: 25300221]
8. Kessler HS. Simple interventions to improve healthy eating behaviors in the school cafeteria. Nutr Rev 2016 Mar;74(3):198-209 [FREE Full text] [doi: 10.1093/nutrit/nuv109] [Medline: 26874753]
9. Bell AC, Swinburn BA. What are the key food groups to target for preventing obesity and improving nutrition in schools? Eur J Clin Nutr 2004 Feb;58(2):258-263 [FREE Full text] [doi: 10.1038/sj.ejcn.1601775] [Medline: 14749745]
10. Evans CEL, Greenwood DC, Thomas JD, Cade JE. A cross-sectional survey of children's packed lunches in the UK: food- and nutrient-based results. J Epidemiol Community Health 2010 Nov;64(11):977-983. [doi: 10.1136/jech.2008.085977] [Medline: 20089755]
11. Dresler-Hawke E, Whitehead D, Coad J. What are New Zealand children eating at school? A content analysis of `consumed versus unconsumed' food groups in a lunch-box survey. Health Education Journal 2009 Mar;68(1):3-13. [doi: 10.1177/0017896908100444]
12. Sabinsky MS, Toft U, Sommer HM, Tetens I. Effect of implementing school meals compared with packed lunches on quality of dietary intake among children aged 7-13 years. J Nutr Sci 2019;8:e3 [FREE Full text] [doi: 10.1017/jns.2018.29] [Medline: 30746124]
13. Clinton-McHarg T, Janssen L, Delaney T, Reilly K, Regan T, Nathan N, et al. Availability of food and beverage items on school canteen menus and association with items purchased by children of primary-school age. Public Health Nutr 2018 Oct;21(15):2907-2914. [doi: 10.1017/S1368980018001726] [Medline: 30070193]
14. Sutherland R, Nathan N, Brown A, Yoong S, Reynolds R, Walton A, et al. A cross-sectional study to determine the energy density and nutritional quality of primary-school children's lunchboxes. Public Health Nutr 2020 Apr;23(6):1108-1116. [doi: 10.1017/S1368980019003379] [Medline: 31969199]
15. Sanigorski AM, Bell AC, Kremer PJ, Swinburn BA. Lunchbox contents of Australian school children: room for improvement. Eur J Clin Nutr 2005 Nov;59(11):1310-1316 [FREE Full text] [doi: 10.1038/sj.ejcn.1602244] [Medline: 16034359]
16. Sutherland R, Nathan N, Brown A, Yoong S, Finch M, Lecathelinais C, et al. A randomized controlled trial to assess the potential efficacy, feasibility and acceptability of an m-health intervention targeting parents of school aged children to improve the nutritional quality of foods packed in the lunchbox 'SWAP IT'. Int J Behav Nutr Phys Act 2019 Jul 02;16(1):54 [FREE Full text] [doi: 10.1186/s12966-019-0812-7] [Medline: 31266506]
17. Evans CEL, Melia KE, Rippin HL, Hancock N, Cade J. A repeated cross-sectional survey assessing changes in diet and nutrient quality of English primary school children's packed lunches between 2006 and 2016. BMJ Open 2020 Jan 13;10(1):e029688 [FREE Full text] [doi: 10.1136/bmjopen-2019-029688] [Medline: 31932386]
18. Everitt T, Engler-Stringer R, Martin W, Vatanparast H. Comparing diet quality of school meals versus food brought from home. Can J Diet Pract Res 2020 Jun 04:1-7. [doi: 10.3148/cjdpr-2020-013] [Medline: 32495636]
19. Hur I, Burgess-Champoux T, Reicks M. Higher quality intake from school lunch meals compared with bagged lunches. ICAN: Infant, Child, & Adolescent Nutrition 2011 Mar 29;3(2):70-75. [doi: 10.1177/1941406411399124]
20. Nathan N, Janssen L, Sutherland R, Hodder RK, Evans CEL, Booth D, et al. The effectiveness of lunchbox interventions on improving the foods and beverages packed and consumed by children at centre-based care or school: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Apr 29;16(1):38 [FREE Full text] [doi: 10.1186/s12966-019-0798-1] [Medline: 31036038]

21. Hammersley ML, Jones RA, Okely AD. Parent-focused childhood and adolescent overweight and obesity eHealth interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jul 21;18(7):e203 [FREE Full text] [doi: 10.2196/jmir.5893] [Medline: 27443862]
22. Zarnowiecki D, Mauch CE, Middleton G, Matwiejczyk L, Watson WL, Dibbs J, et al. A systematic evaluation of digital nutrition promotion websites and apps for supporting parents to influence children's nutrition. Int J Behav Nutr Phys Act 2020 Feb 10;17(1):17 [FREE Full text] [doi: 10.1186/s12966-020-0915-1] [Medline: 32041640]
23. Australian Dietary Guidelines. National Health Medical Research Council. 2013. URL: https://www.nhmrc.gov.au/adg [accessed 2021-05-01]
24. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [Medline: 27777223]
25. Sutherland R, Brown A, Nathan N, Janssen L, Reynolds R, Walton A, et al. Protocol for an effectiveness- implementation hybrid trial to assess the effectiveness and cost-effectiveness of an m-health intervention to decrease the consumption of discretionary foods packed in school lunchboxes: the 'SWAP IT' trial. BMC Public Health 2019 Nov 12;19(1):1510 [FREE Full text] [doi: 10.1186/s12889-019-7725-x] [Medline: 31718597]
26. 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]
27. Wolfenden L, Reilly K, Kingsland M, Grady A, Williams CM, Nathan N, et al. Identifying opportunities to develop the science of implementation for community-based non-communicable disease prevention: A review of implementation trials. Prev Med 2019 Jan;118:279-285. [doi: 10.1016/j.ypmed.2018.11.014] [Medline: 30468794]
28. Blom-Hoffman J, Leff SS, Franko DL, Weinstein E, Beakley K, Power TJ. Consent procedures and participation rates in school-based intervention and prevention research: using a multi-component, partnership-based approach to recruit participants. School Ment Health 2009 Mar 01;1(1):3-15 [FREE Full text] [doi: 10.1007/s12310-008-9000-7] [Medline: 19834586]
29. Pechey R, Monsivais P. Socioeconomic inequalities in the healthiness of food choices: Exploring the contributions of food expenditures. Prev Med 2016 Jul;88:203-209 [FREE Full text] [doi: 10.1016/j.ypmed.2016.04.012] [Medline: 27095324]
30. Census of Population and Housing: Socio-Economic Indexes for Australia (SEIFA). Australian Bureau of Statistics (ABS). Canberra; 2016. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012016?OpenDocument [accessed 2021-05-01]
31. Michie S, van SMM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
32. Reynolds R, Sutherland R, Nathan N, Janssen L, Lecathelinais C, Reilly K, et al. Feasibility and principal acceptability of school-based mobile communication applications to disseminate healthy lunchbox messages to parents. Health Promot J Austr 2019 Jan;30(1):108-113. [doi: 10.1002/hpja.57] [Medline: 29719083]
33. Nutrition in schools policy. NSW Department of Education. 2011. URL: https://policies.education.nsw.gov.au/policy-library/policies/nutrition-in-schools-policy [accessed 2021-05-01]
34. Brown A, Sutherland R, Janssen L, Hudson N, Chooi A, Reynolds R, et al. Enhancing the potential effects of text messages delivered via an m-health intervention to improve packing of healthy school lunchboxes. Public Health Nutr 2020 Oct 14:1-10. [doi: 10.1017/s1368980020003997]
35. Kremer PJ, Bell AC, Swinburn BA. Calibration and reliability of a school food checklist: a new tool for assessing school food and beverage consumption. Asia Pac J Clin Nutr 2006;15(4):465-473 [FREE Full text] [Medline: 17077061]
36. Mitchell SA, Miles CL, Brennan L, Matthews J. Reliability of the School Food Checklist for in-school audits and photograph analysis of children's packed lunches. J Hum Nutr Diet 2010 Feb;23(1):48-53. [doi: 10.1111/j.1365-277X.2009.00996.x] [Medline: 19788706]
37. Hardy L, Mihrshahi S, Drayton B, Bauman A. NSW schools physical activity and nutrition survey (SPANS) 2015: full report. NSW Department of Health. Sydney; 2016. URL: https://www.health.nsw.gov.au/heal/Publications/spans-2015-full-report.PDF [accessed 2021-05-01]
38. Miles CL, Matthews J, Brennan L, Mitchell S. Changes in the content of children's school lunches across the school week. Health Promot J Austr 2010 Dec;21(3):196-201 [FREE Full text] [doi: 10.1071/he10196] [Medline: 21118066]
39. Nathan N, Wolfenden L, Butler M, Bell AC, Wyse R, Campbell E, et al. Vegetable and fruit breaks in Australian primary schools: prevalence, attitudes, barriers and implementation strategies. Health Educ Res 2011 Aug;26(4):722-731. [doi: 10.1093/her/cyr033] [Medline: 21571836]
40. Sorhaindo A, Feinstein L. Wider Benefits of Learning Research Report no. 18. What Is the Relationship Between Child Nutrition and School Outcomes?. London: Centre for Research on the Wider Benefits of Learning; 2006.
41. Fredricks J, Blumenfeld P, Friedel J. School engagement. In: Moore KA, Lippman LH, editors. What Do Children Need to Flourish? Conceptualising and Measuring Indicators of Positive Development. New York: Springer Science and Business Media; 2005:305-321.

42. Rao M, Afshin A, Singh G, Mozaffarian D. Do healthier foods and diet patterns cost more than less healthy options? A systematic review and meta-analysis. BMJ Open 2013 Dec 05;3(12):e004277 [FREE Full text] [doi: 10.1136/bmjopen-2013-004277] [Medline: 24309174]
43. Australian statistical geography standard (ASGS): volume 5 - remoteness structure. Australian Bureau of Statistics (ABS). Canberra; 2016. URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.005 [accessed 2021-05-01]
44. Cochrane T, Davey R, de Castella FR. Estimates of the energy deficit required to reverse the trend in childhood obesity in Australian schoolchildren. Aust N Z J Public Health 2016 Feb;40(1):62-67 [FREE Full text] [doi: 10.1111/1753-6405.12474] [Medline: 26561382]
45. Wang YC, Orleans CT, Gortmaker SL. Reaching the healthy people goals for reducing childhood obesity: closing the energy gap. Am J Prev Med 2012 May;42(5):437-444. [doi: 10.1016/j.amepre.2012.01.018] [Medline: 22516482]
46. Practical guidance for scaling up health service innovations. World Health Organization. Geneva; 2009. URL: https://www.who.int/reproductivehealth/publications/strategic_approach/9789241598521/en/ [accessed 2021-05-01]
47. Hardiker NR, Grant MJ. Barriers and facilitators that affect public engagement with eHealth services. Stud Health Technol Inform 2010;160(Pt 1):13-17. [Medline: 20841641]
48. Delaney T, Wyse R, Yoong SL, Sutherland R, Wiggers J, Ball K, et al. Cluster randomized controlled trial of a consumer behavior intervention to improve healthy food purchases from online canteens. Am J Clin Nutr 2017 Nov;106(5):1311-1320. [doi: 10.3945/ajcn.117.158329] [Medline: 28971849]
49. Lal A, Peeters A, Brown V, Nguyen P, Tran HNQ, Nguyen T, et al. The modelled population obesity-related health benefits of reducing consumption of discretionary foods in Australia. Nutrients 2020 Feb 28;12(3):649-663 [FREE Full text] [doi: 10.3390/nu12030649] [Medline: 32121199]
50. Minero E. Parent engagement in the digital age.: George Lucas Educational Foundation, Edutopia; 2017 Jul 1. URL: https://www.edutopia.org/article/parent-engagement-digital-age [accessed 2020-07-01]
51. Cappellini B, Harman V, Parsons E. Unpacking the lunchbox: biopedagogies, mothering and social class. Sociol Health Illn 2018 Sep 19;40(7):1200-1214. [doi: 10.1111/1467-9566.12751] [Medline: 29916555]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
NSW: New South Wales
SEIFA: Socio-Economic Index for Areas
SFC: School Food Checklist
WHO: World Health Organization

Edited by G Eysenbach; submitted 25.10.20; peer-reviewed by S Browne, P Ducrot; comments to author 03.12.20; revised version received 20.12.20; accepted 20.01.21; published 24.06.21.

Please cite as:
Sutherland R, Brown A, Nathan N, Yoong S, Janssen L, Chooi A, Hudson N, Wiggers J, Kerr N, Evans N, Gillham K, Oldmeadow C, Searles A, Reeves P, Davies M, Reilly K, Cohen B, Wolfenden L
A Multicomponent mHealth-Based Intervention (SWAP IT) to Decrease the Consumption of Discretionary Foods Packed in School Lunchboxes: Type I Effectiveness–Implementation Hybrid Cluster Randomized Controlled Trial
J Med Internet Res 2021;23(6):e25256
URL: https://www.jmir.org/2021/6/e25256/
doi:10.2196/25256
PMID:34185013

Original Paper

Receptiveness and Responsiveness Toward Using Social Media for Safe Firearm Storage Outreach: Mixed Methods Study

Esther Lam[1], MPH; Megan Moreno[2], MD, MSED, MPH; Elizabeth Bennett[3], MPH, MCHES; Ali Rowhani-Rahbar[1], MD, MPH, PhD

[1]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, United States

[2]School of Medicine and Public Health, University of Wisconsin, Madison, WI, United States

[3]Seattle Children's Hospital, Seattle, WA, United States

Corresponding Author:
Esther Lam, MPH
Department of Epidemiology
School of Public Health
University of Washington
UW Box #351619
Seattle, WA, 98195
United States
Phone: 1 4088901256
Email: estherwmlam@gmail.com

Abstract

Background: Childhood and adolescent firearm injury and death rates have increased over the past decade and remain major public health concerns in the United States. Safe firearm storage has proven to be an effective measure to prevent firearm injury and death among youth. Social media has been used as an avenue to promote safe firearm storage, but perceptions of this tool remain unknown.

Objective: The aim of this study was to determine receptiveness and responsiveness in promoting firearm lock box and trigger lock giveaway events on social media, and to describe the characteristics of participants who learned of these events through social media.

Methods: We performed a mixed methods study combining a content analysis of Facebook event post comments, quantitative analysis of positive and negative feedback on social media, and a descriptive analysis of event participant characteristics. Through a qualitative content analysis approach, we thematically coded comments from each event's social media page posting. Interrater reliability and κ statistics were calculated. We calculated the prevalence of positive and negative feedback data. Further, we calculated descriptive statistics for demographic characteristics gathered from day-of-event intake surveys. Differences between collected measures were analyzed with χ^2 and *t* tests according to how the participant found out about the event (social media vs other means). Using concurrent analysis, we synthesized the results from both the qualitative and quantitative aims.

Results: Through qualitative content analysis, 414 comments from 13 events were coded. Seven themes emerged through the comment coding process with the most common being "positive receptiveness" (294/414, 71.0%). From quantitative analysis of the social media content, we found higher levels of positive feedback compared to negative feedback. The average number of event post "likes" was 1271.3 per event, whereas the average count in which "hide post" was clicked was 72.3 times per event. Overall, 35.9% (1457/4054) of participants found out about the event through social media. The participants who learned about the event through social media were on average significantly younger than those who learned about the event through other means (–6.4 years, 95% CI –5.5 to –7.3). Among the group that learned of the event through social media, 43.9% (629/1433) identified as female, whereas 35.5% (860/2420) identified as female among the group that learned of the event through other means.

Conclusions: There was overall positive receptiveness and responsiveness toward firearm lock box and trigger lock giveaway events when promoted on social media. Compared with other promotional tools, social media has the ability to reach those who are younger and those who identify as female. Future studies should extend this research to determine whether there is a difference between rural and urban settings, and consider other social media platforms in the analysis.

(J Med Internet Res 2021;23(6):e24458) doi:10.2196/24458

KEYWORDS

firearm storage; gun safety; public health outreach; social media; mixed methods; family

Introduction

Background

Firearm injury is a major public health concern in the United States. Among children under 17 years of age, approximately 1300 die from and 5790 are treated for gunshot wounds each year [1]. Firearm suicide rates among this population have been increasing over the past few years [1,2] along with increasing rural-urban divides [3-5]. Based on a national study of individuals younger than 20 years, the incidence of firearm injury hospitalization was the highest among 15 to 19-year-olds residing in urban settings [4]. However, hospitalization rates were higher in rural areas than in urban areas for those in the 5-9 and 10-14–year age groups [4].

Safe storage of firearms is associated with reduced risk of unintentional and intentional self-inflicted firearm injury among youth [6]. Examples of safe firearm storage practices include storing the firearm in a locked location such as a gun safe or lockbox, or with a trigger lock, and storing it unloaded with the ammunition stored separately [6]. A study conducted in the United States found that up to 65% of all surveyed high school seniors in a midwestern state had access to at least one firearm in their household [7]. An evaluation with a nationally representative sample found that approximately 1 in every 3 adolescents reported living in a home with firearm access [8]. Moreover, 1 in every 5 firearm-owning households with children store their firearms in the least safe manner (ie, loaded and unlocked) [9].

In Washington state, 34% of households reported having a gun at home, and only 38% of those reported storing their guns locked and unloaded [10]. In a survey of 2956 participants at safe firearm storage events in Washington state, 40.1% indicated that they stored at least one firearm unlocked [11].

Washington state passed Initiative Measure No. 1639 in 2018 [12], which requires a semiautomatic rifle purchaser to provide proof of completion of a recognized firearm safety training program that includes secure gun storage education [12]. Additionally, under this initiative, a person who leaves a firearm in a place where a prohibited person (ie, a child) could potentially gain access to the firearm will be charged with community endangerment if the prohibited person gains access to the firearm [12].

Safe Firearm Storage Giveaway Outreach Events

Prior to the passing of Initiative Measure No. 1639, Seattle Children’s Hospital began the “Safe Firearm Storage Giveaway Outreach Events” (SFSGOE) program. This program sought to reach both urban and rural communities. The primary mission of these events was to prevent firearm injuries by creating safe household environments for children, their families, and their communities, and provide free firearm lock boxes and trigger locks supplemented with education and device demonstrations. The events were focused on parents living in households with children. This was a community-based effort involving multiple stakeholders, including other hospitals; public health agencies; coalitions aimed at advocating for safe environments for children; and sporting goods stores that sell firearms, lock boxes, and trigger locks. These efforts have proven to be effective in changing behavior related to safe firearm storage practices. An evaluation survey performed 4-6 weeks after the event found that a significantly greater proportion of households with children and adolescents reported having stored and locked their firearms safely [13]. Flyers, newspaper advertisements, radio, television, sharing through word of mouth, and posting on social media were all used to promote these events. Social marketing, designed to influence health behaviors, prevent injuries, and contribute to communities on a large scale, was used as a framework for promotion and messaging [14,15].

Social media offers a versatile and unique tool for implementing a social marketing approach to increase the number of individuals and communities reached. Accordingly, marketing using social media has become a growing trend in many fields. However, there has been little research performed to understand whether or not social media promotion of topics such as firearm use and storage practices is effective in targeting the appropriate audience and yielding acceptance. Skepticism may arise regarding whether social media platforms are the appropriate virtual venues for potentially controversial conversations, which can yield negative reactions. Therefore, the purpose of this study was to understand the individual- and community-level responses toward social media as an educational and marketing tool for promoting safe firearm storage giveaway events. We also sought to evaluate the attitudes and conversations that emerged through the social media marketing event page posts.

Study Aims

There were three aims of this study: (1) understand the receptiveness of using social media to promote safe firearm storage in the continuum of rural to urban areas; (2) assess the individual-level responsiveness toward the SFSGOE hosted by Seattle Children’s Hospital; and (3) compare the characteristics of event participants who learned of the event via social media with those of participants who learned of the event through other sources.

Methods

Study Design

We used a mixed methods design, incorporating both qualitative and quantitative approaches, which allows for gaining a deeper understanding of pragmatic implementation in addition to increasing reproducibility [16]. Furthermore, other recent studies with the aim of understanding the intersection between social media and health behaviors or conditions have also used this mixed methods approach to allow for a comprehensive review of all metrics and data available on the social media platform of interest [17,18].

For this mixed methods approach, we utilized a convergent design. Since both elements of this study pull from existing data, quantitative analysis and qualitative evaluation occurred concurrently, and the results were then merged to supplement each other upon synthesis.

Conceptualized by our research team for the purpose of this study, receptiveness was defined as the attitudes toward the event and responsiveness was defined as the quantifiable interactions on the event page posts.

Assessing Rurality Versus Urbanity

The United States Department of Agriculture Economic Research Service Rural-Urban Continuum Codes (RUCC) from 2013 were used to categorize the locations of these events [19]. The 2013 edition was the most recent RUCC available at the time of analysis as these codes are updated every 10 years [19]. Codes are assigned to counties, ranging from 1 to 9 based on population size and proximity to a metropolitan area [19]. Urban counties are coded between 1 and 3; code 1 is given to a city in a large-sized urban county, code 2 is given to a city in a medium-sized urban county, and code 3 is given to a city in a small semiurban county [19]. Nonmetropolitan counties are coded from 4 to 9, with 4 being a city in a larger semirural county and 9 being a city in a small completely rural county [19]. Table 1 lists the SFSGOE locations and their assigned RUCCs.

Table 1. Locations and times of events that were evaluated along with their rural and urban codes (RUCC).

Month and year of event	City	County	RUCC
December, 2014	Seattle	King	1
January, 2015	Fife	Pierce	1
April, 2015	Kennewick (Tri-Cities)	Franklin	2
June, 2015	Monroe	Snohomish	1
October, 2015	Tacoma	Pierce	1
November, 2015	Kirkland	King	1
June, 2016	Toppenish	Yakima	3
July, 2016	Marysville	Snohomish	1
October, 2016	Wenatchee	Chelan	3
February, 2017	Seattle	King	1
May, 2017	Mount Vernon	Skagit	3
June, 2017	Lacey	Thurston	2
October, 2017	Moses Lake	Grant	5
March, 2018	Silverdale	Kitsap	2

Aim 1: Content Analysis to Assess Receptiveness

Overview of Content Analysis

Content analysis was used to evaluate comments generated by Facebook users in response to posts advertising and promoting SFSGOE. For this social media content analysis, a codebook was iteratively created to capture the themes in attitudes and perceptions toward these events presented in the coded comments.

Data Source

We used data from the social media platform, Facebook, for this analysis because Facebook was the primary platform used for promoting the SFSGOE. Seattle Children's Hospital created a Facebook event page, posting details of each event and also shared the event on their main Facebook page. The page's visibility was dependent on the sharing of the event and set reminders innate to Facebook event pages to provide additional promotion.

Data Query

A Facebook event page post query was performed. We searched the Seattle Children's Hospital Facebook posts sharing information about the event and for the Facebook event page. Only comments and metrics linked to the event page that were from Seattle Children's Hospital shared posts were used in the analysis. Upon searching for these retrospective posts, the study team did not interact (ie, like, share, comment) with the posts; all data provided for this analysis were strictly independent of the study team and unaltered.

Data Collection

Data from December 2014 to March 2018 were used for this study. This period included the first event for which Facebook was used to promote the SFSGOE until the most recent Facebook-promoted event at the time of this analysis. All comments linked to the event page posting were deidentified and saved in a password-protected Microsoft Excel spreadsheet. Data from these social media comments were then qualitatively evaluated via exploratory content analysis to understand receptiveness and were quantitatively evaluated to understand responsiveness.

Codebook Development and Coding Validation

An iterative categorization approach was used for this analysis to allow for inductive coding, resulting in a structured system that would allow for reproducible results [20]. Using this inductive coding approach, parent and child codes gradually emerged and were created. Parent codes were the overarching salient themes and child codes stemmed from each of the parent codes, with much more specific attitudes and perspectives [21]. Owing to the finite quantity of comments available for this content analysis, we chose to code all available data to ensure that the themes generated from the sampling would be representative. Considering this to be an exploratory content analysis, we aimed to use the inductive thematic saturation model [22]. Since we did not use a priori conceptualized categories or an existing codebook, this model would allow for generation of themes found to be most appropriate to the generated responses from SFSGOE Facebook event pages [22].

At the time of this analysis, there was no known validated assessment for emoticons; therefore, emoticons were not coded. Only comments in the English language were coded. Each comment was only allowed one code; therefore, for lengthier comments, the most explicit and apparent theme was coded. The comment and the assigned code were recorded in a password-protected Microsoft Excel spreadsheet.

All comments, including replies and comments from Seattle Children's Hospital, were coded. Initial coding was performed by one study team member with prior training in qualitative research methods. Texts from the comments were recorded verbatim, excluding identifiers such as the name of the Facebook user or names of any tagged users.

To evaluate the coding process and to ensure the validity of the codebook, interrater reliability and the κ statistic were calculated. Another study team member with prior training in content analysis, and specifically with this codebook, was given 10% (n= 42) of the comments to code independently. Coding results from this research team member and the first set of coding results were compared. The results indicated a strong level of agreement ($\kappa=0.87$, $P<.001$) supporting the reliability of the coding procedures.

Aim 2: Quantitative Measures and Responsiveness Analysis

Each event allowed for "liking" of the post in addition to commenting. Post "likes" are publicly available data. In addition to these data, we used the SimplyMeasured software system, which tracks deidentified data, including the number of clicks and the "total reach" of the event page post. The "total reach" was defined as the total number of unique individuals who saw the post during the "reporting period." The "reporting period" was the duration when the post was publicly available. We also qualified the data system in terms of "negative feedback." Extracted "negative feedback" data displayed the number of individuals who selectively chose to hide the post, report the post as spam, hide all future posts from the same posting account, and those who unliked the page of the account that posted these events.

For each event, prevalence was calculated using the count of each measure as the numerator and the "total reach" number as the denominator.

Aim 3: Survey

Data Collection, Measures, and Inclusion Criteria

At each SFSGOE, as part of the intake and prior to receipt of a trigger lock or lock box, each household was given a paper survey to complete. This survey included demographic questions, current firearm practice questions, questions related to how the household found out about the event, and their intention to use the giveaway device in the next week. Only those who identified as 18 years of age or older were provided the equipment. No questions on the survey were mandatory. An individual's receipt of a safe firearm storage device was not dependent on completion of the survey. The submission of a survey and release form were used as a proxy to quantify the attendance of these events. The outreach event staff then entered the survey data into a secure database in REDCap [23,24]. Data from the database were then extracted and quantitatively analyzed. Only individual-level responses deemed to be "complete" by the data importer were used.

Statistical Analysis

The χ^2 test was used to examine whether there was a significant difference in distribution of the categorical variables (gender, firearm storage practices, and intention to use in the next week) based on how the individual found out about the event (social media vs other means). The *t* test was used to examine if there was a significant difference in mean age according to how individuals found out about the event. A *P* value less than .05 was used as the cutoff for significance.

Ethical Approval

The content analysis portion of this study was exempt from Institutional Review Board review because it only used publicly accessible and deidentified data from Facebook. The survey used in this study was approved by the Institutional Review Board of Seattle Children's Hospital, Seattle, Washington.

Results

Aim 1: Content Analysis

A total of 418 comments from 13 event page posts were publicly available during the timeframe of interest. A total of 414 comments were coded; 4 comments were not evaluated because they only used emoticons or were not written in English. From the 15 events that occurred within the timeframe, only 13 resulted in publicly available posts during the data search and extraction. On average, event posts had 32 comments (range 0-131). Of the 13 event page posts evaluated, eight were from areas that were coded as a city in a large urban county (RUCC=1), one was categorized as a city in a medium-sized urban city (RUCC=2), three were categorized as a city in a small-sized semiurban county (RUCC=3), and one was categorized as a city in a medium-sized semirural county (RUCC=5).

A total of seven parent codes were developed during the iterative coding process: (1) Positive Receptiveness, (2) Negative Receptiveness, (3) Postevent Commenting, (4) Advocacy Against Firearm Storage and/or Firearm Control, (5) Advocacy for Firearm Storage and/or Firearm Control, (6) Commenting to Moderate, and (7) Comments by the Host. Stemming from these parent codes, 21 child codes were created. The coding schemas at the end of the inductive coding process along with representative examples of comments are presented in Textbox 1.

The most common parent code was Positive Receptiveness, accounting for 71.0% (294/414) of all comments coded. Positive Receptiveness is defined as a comment showing a positive attitude toward the event, including expressing excitement, gratitude, or interest explicitly, or sharing the event by tagging friends. Expressing excitement or gratitude toward the event accounted for 10.9% (45/414) of all comments. Of all comments coded, 40.3% (167/414) involved Facebook users tagging other Facebook users, which may be an implicit notion of sharing the event post. Among all codes, 6.3% (26/414) of all comments were coded as the Facebook user showing interest for the event, but could not make it to the event due to the distance and timing of the event.

Negative Receptiveness made up 3.6% (15/414) of all comments. Of all comments, 2.9% (12/414) involved individuals expressing that they did not understand why these events were necessary or that they did not think that lock boxes work. Overall, 0.7% (3/414) of individuals expressed doubtful sentiments about the potential success of the event. Postevent commenting made up 0.7% (3/414) of all codes. This included participants going back to the event post and commenting about their experience at the event. Most of these comments were factual experiences, including how long the wait was to receive their giveaway device or what type of firearm storage device they received at the event.

Advocacy Against Firearm Storage and/or Firearm Control made up 3.9% (16/414) of total comments. These comments included sentiments that children should be trained to use firearms and that there should be much more liberal firearm ownership. Advocacy for Firearm Storage and/or Firearm Control made up 7.0% (29/414) of the coded comments. These comments included perspectives that children should be well protected from firearms and there also were some comments that expressed sentiments of wanting more firearm control.

The parent code Commenting to Moderate was used to denote a case in which a Facebook user commented during or subsequent to a political or controversial discussion. We defined this as moderating, as this type of comment would aim at drawing attention back to the goal of the event, alleviating the tone created by controversial comments; 2.2% (9/414) of the coded comments were coded as such. This emergence of commenting to moderate follows with the emergence of “moderators.” For the purpose of this study, a moderator was defined as a Facebook user, not associated with Seattle Children’s Hospital, who facilitates the conversation and draws other Facebook users back to the goal of the event (ie, to promote safe firearm storage practices). These moderating comments were not from the Seattle Children’s Hospital account and there were no known intentions from the hospital to have users take up this role. This was a role that emerged naturally through discussions from Facebook user interactions on the event page posts.

Comments by the Host as a parent code was defined as any comment posted by Seattle Children’s Hospital to answer any inquiries from a comment or to thank commenting users for their support in some cases; 9.4% (39/414) of all comments were Comments by the Host.

Comments that were not relevant to the topic of the event were not coded with any parent or child code and were marked as Spam (2.2%, 9/414).

Textbox 1. Inductively created codebook and representative comments from evaluation of event page posts.

Positive Receptiveness

Any comment showing a positive attitude toward the event, including expressing excitement, gratitude, or interest explicitly or sharing the event by tagging friends.

- Sharing the event by tagging another Facebook user
- Asking for more of these events or suggesting other locations for these events
- Showing appreciation/gratitude toward Seattle Children's Hospital for hosting these events
- Showing excitement for the event, positive exclamation, or emotion for the event (eg, "Yay!" "Awesome!")
- Calling for others to go to the event by sharing more details about the event
- Expressing interest for the event but also saying that they cannot go to the event
- Saying that they will go to the event
- Positive comments about lock boxes or trigger locks
- Asking for more information about the event

Example: "Thank you Seattle Children's and Outdoor Emporium. This was just what my family needed" [Facebook user]

Negative Receptiveness

Any comment showing a negative attitude toward the event, including expressing doubt for either the event or a safe firearm storage device

- Does not understand why these events are necessary, or why lock boxes or trigger locks are necessary
- Doubtful about how these events will be successful
- Shared past negative experience with a lock box or trigger lock

Example: "This is a nice give away but I doubt there will be enough boxes...there will be plenty of locks. Every gun has a lock sold with it. And will never be used by anyone with brains" [Facebook user]

Postevent Commenting

Any comment shared after the event by an event participant and comments related to their experience at the event

- Shared facts about experience at the event without a sense of positivity or negativity

Example: "After they ran out of Bulldog brand they started handing out Fortress brand lockboxes" [Facebook user]

Advocacy Against Firearm Storage and/or Firearm Control

Any comment displaying opposition against safe firearm storage (including lock boxes and trigger locks)

- Shared a story anecdotally or factually that children can handle guns well, or advocating that children learn how to handle guns
- Words of disagreement or discourse against firearm safety
- Advocating for liberal gun ownership

Example: "Problem is in a home invasion you don't have time to unlock your safe gun box. You lose, that's what happened to my brother" [Facebook user]

Advocacy for Firearm Storage and/or Firearm Control

Any comment displaying support for safe firearm storage (including lock boxes and trigger locks)

- Advocating that children should be well protected from guns
- Words of agreement or discourse relating to firearm safety
- Bringing up banning gun ownership

Example: "But does every other kid who comes in your home? I'm all for firearm education, I went through it all when I was a kid, but fewer kids these days are learning from real educators and more from cartoons and movies. Not trying to pick a fight, just posing a question. My 4 year old play shoots (much against my wishes) and my almost 2 year old has picked up on it. The older one comprehends the severity of death, but the baby has no idea. Lock boxes and trigger locks isn't to protect the ones who understand, it's to protect the ones who don't" [Facebook user]

Commenting to Moderate

Any comment from a Facebook user that draws attention back to the mission of the goal of the event

- Moderates a heated discussion and draws attention back to the mission of the event

Example: “XXX, I’d expect an anti-gun advocate to be more collected. if you are truly trying to win over any readers to be anti-gun advocates perhaps leave out the personal slams. I do hope you read your messages to XXX, and sincerely apologize to her. In my humble opinion it was completely unwarranted. Have a lovely evening.” [Facebook user]

Comments by Host

Any comment from Seattle Children’s Hospital

- Seattle Children’s Hospital outreach comment or response to an inquiry

Example: “We have a couple hundred items to give away, XXX. If you can’t come to the giveaway Saturday, visit www.lokitup.org for retailers in King County who are offering a discount on select storage devices or lock boxes through December 2014.” [Facebook user]

Spam

Anything unrelated to the topic of firearm safety, Seattle Children’s Hospital, or the event

Aim 2: Quantitative Analysis for Responsiveness

Prevalence and Reach

Among the 13 events for which comments were coded, 4 did not yield data from SimplyMeasured to allow for positive and negative feedback prevalence to be calculated. Among the 9 events with this type of data, the total reach on Facebook was 491,155 (range 10,536-187,072).

Positive Feedback

The average number of likes was 1271.3 per event, with the highest of 13.63 per 1000 engagements. Engagements were defined as the average number of likes, comments, and shares on all posts published during the reporting period. The average number of clicks per event was 5183.6 for all events. The range of clicks was 18.0-76.5 per 1000 engagements.

Negative Feedback

The average count of times in which “hide post” was clicked on the event posts was 72.3, equating to 1.3 counts per 1000 engagements. The highest prevalence of such clicks was 5.8 per 1000 engagements and the lowest was 0.41 counts per 1000 engagements. There were no posts reported as spam. The average count of “hiding all future posts” was 8.6, equating to 0.16 counts per 1000 engagements. Among all events, the highest prevalence of “hiding all future posts” was 0.88 per 1000 engagements, whereas the lowest was 0.1 per 1000 engagements. There were only 2 counts of “unliking” the page among the 9 events, and both instances happened on the same event post.

Aim 3: Quantitative Survey Data

Although there were 15 events that occurred during the timeframe of interest, two events did not use the intake survey. Owing to lack of data for these two events, these two event locations were not included in this portion of the analysis. One of the locations missing survey data was for an event that was held in a city in a large urban county (RUCC=1). The second event without survey data was located in a city in a small semiurban county (RUCC =3).

There were two versions of the paper survey: one for the first 3 events and one for the subsequent 10 events. The question regarding location of residence was assessed in the second version of the survey, but not in the first. Owing to the lack of data for the first three events and high levels of missing data, this question was not incorporated in the analysis. There were overlapping questions between the two surveys (see Multimedia Appendix 1 for the sample survey). The questions that overlapped across the two versions of the survey were used. For the purposes of this analysis, similar question answers were aggregated.

Only data with completed event location information and marked as “complete” by the data importer were used for this evaluation. Among the 5512 survey records found in the survey database, only 4054 (81.7%) were marked as “complete” by the data importer, and included complete data on how the participant found out about the event and the location of the event.

Based on the question inquiring how the participant found out about the giveaway event, 1457 (35.9%) indicated some form of social media. Alternatively, 2597 (64.1%) participants indicated that they learned about the event through the newspaper, a friend, the radio, a flyer, word of mouth, working at the store, came to the store at the time of the event, or through other means. See Table 2 for more details on the distribution of how participants learned of these events.

Table 2. Proportion of participants' responses as to how they learned about Safe Firearm Storage Giveaway Outreach Events (SFSGOE) 2014-2018 based on the intake survey.

How did you learn of these SFSGOE?[a]	Responses, n (%)[b]
Social media[c]	1457 (34.6)
Newspaper	383 (9.1)
Friend	329 (7.8)
Radio[d]	41 (1.0)
(Event) Flyer	261 (6.2)
Word of mouth[e]	638 (15.1)
Work at the store[e]	37 (0.9)
Came to the store when the event was occurring[e]	411 (9.7)
Other	660 (15.7)

[a] Participants were asked to "mark all that apply"; therefore, the total of these responses (N=4217) exceeds the number of total participants.

[b] Based on the total number of responses.

[c] With two versions of the survey, the version used for the last 10 events included distinguishing whether the participant found out about the event through Facebook or Twitter. These data were aggregated given the small number of participants who reported Twitter (<1.0%).

[d] With two versions of the survey, the version used for the first three events included radio as a potential response that was not included as an option for respondents to the second version of the survey. This option may have been answered as "other" for participants who used the second version of the survey.

[e] With two versions of the survey, the version used for the last 10 events included these responses that were not included as options for participants who responded to the earlier version of the survey. These options may have been answered as "other" for participants who used the earlier version of the survey.

With respect to the distribution of participant data across locations, 93.2% (3778/4054) of participants with a completed survey were attendees of an urban-located event. Out of all participants, 6.8% (276/4054) with a completed survey were attendees of a semirural-located event. The number of completed surveys for these events ranged from 161 to 406. The event with the least number of completed surveys was the event held in November of 2015 (Kirkland) and that with the largest number of completed surveys was the event held in January of 2015 (Fife). Due to the skew of data toward urban event participants, we were unable to find results that were statistically significant in comparing how participants found out about the event and the location of the event (*P*=.08).

Table 3 compares the characteristics of participants who learned of the event through social media with those of participants who learned about the event through other means. On average, participants who learned about the event via social media were 6.4 years (95% CI 5.5-7.3) younger than those who learned about the event via other means.

In both groups, more participants identified as male. A slightly lower proportion of those who learned about the event via social media indicated that they were not current users of a gun storage/safety device (*P*=.30).

A high intention to use was identified based on survey responses for the giveaway firearm storage device for all event participants, regardless of how they found out about the events; however, the intention to use was significantly higher among the social media group than the nonsocial media group (*P*=.001).

Table 3. Characteristics of participants who found out about Safe Firearm Storage Giveaway Outreach Events (2014-2018) through social media or other means.

Participant characteristic/survey response	Social media (n=1457)	Other means (n=2597)
Age (years), mean (SD)	38.8 (11.8)	45.2 (15.5)
Age category (years), n (%)[a]		
18-29	311 (22.8)	410 (18.0)
30-39	500 (36.7)	539 (23.7)
40-49	286 (21.0)	422 (18.5)
50-59	170 (12.5)	426 (18.7)
60-69	84 (6.2)	338 (14.8)
≥70	11 (0.8)	143 (6.3)
Gender, n (%)[b]		
Male	801 (55.9)	1553 (64.2)
Female	629 (43.9)	860 (35.5)
Other	3 (0.2)	7 (0.3)
Currently use gun storage device[c], n (%)		
Yes (gun safe, gun lock box, trigger lock, cable lock, other)	1015 (70.5)	1698 (68.7)
No	424 (29.5)	772 (31.3)
Currently store gun(s) loaded[d], n (%)		
Yes, all of them	428 (30.0)	568 (23.4)
Yes, some of them (some are and some aren't)	320 (22.3)	536 (22.1)
None of them	578 (40.3)	1078 (44.4)
Not sure	15 (1.0)	36 (1.5)
Does not apply to my home	93 (6.5)	211 (8.7)
Intend to use gun safety device in the next week[e], n (%)		
Yes	1331 (92.2)	2201 (89.1)
No	39 (2.7)	119 (4.8)
Unsure/not sure	73 (5.1)	151 (6.1)
RUCC[f] based on location of event[g], n (%)		
1: urban (large county)	783 (53.7)	1291 (49.7)
2: urban (medium county)	412 (28.3)	701 (27.0)
3: urban (small county)	149 (10.2)	442 (17.0)
4: semirural (large county)	0 (0.0)	0 (0.0)
5: semirural (medium-large county)	113 (7.8)	163 (6.3)

[a] Due to missing answers, 1362 and 2278 participants responded to this question in the social media and nonsocial media group, respectively.

[b] Due to missing answers, 1433 and 2420 participants responded to this question in the social media and nonsocial media group, respectively.

[c] Due to missing answers, 1439 and 2470 participants responded to this question in the social media and nonsocial media group, respectively. For this question, participants were asked to "mark all that apply"; therefore, the total number of recorded responses (N=4925) is larger than the number of participants. The option "other" was added to the second version of the survey. These data represent aggregated responses for indication of use of any of the described firearm storage device and "other."

[d] Due to missing answers, 1434 and 2429 participants responded to this question in the social media and nonsocial media group, respectively. This question and the answer options were different for the two versions of the survey. The second version added the option of "not sure" and "does not apply to my home."

[e] Due to missing answers, 1443 and 2471 participants responded to this question in the social media and nonsocial media group, respectively.

[f] RUCC: United States Department of Agriculture Economic Research Service Rural-Urban Continuum; codes from 2013 were used to categorize locations (refer to Table 1). There were no events held and included in this analysis for RUCC 6, 7, 8 and 9.

[g] Due to missing answers, 1457 and 2597 participants responded to this question in the social media and nonsocial media group, respectively.

Discussion

Principal Findings

To our knowledge, this study is one of the first to investigate the receptiveness and responsiveness of using social media as a marketing tool for a safe firearm storage giveaway intervention. Prior to this study, it was unknown whether or not there would be receptiveness and responsiveness in using social media to market these events. Much of this uncertainty stemmed from acknowledging the controversial nature of this topic and the need to assess whether there would be acceptance of using this tool. Additionally, if there was acceptance, it was unclear whether the level of acceptance would differ between rural and urban areas.

Including the two terms of "receptiveness" and "responsiveness" to conceptualize the aims of the study helped with conciseness, and encapsulated the goals of health and social media research, especially with the public-facing metrics and data available on Facebook. As these terms are generally defined, they can be expanded upon and used by other social media researchers when trying to understand the acceptability and reach of health promotion campaigns for other social media platforms as well.

We found high levels of acceptability and positive receptiveness toward these social media posts through user comments. The content analysis showed a higher proportion of positive than negative comments. Facebook generated an avenue for users to not only find out about the event but also to share the event through easily "tagging" other Facebook users in the comment. Individuals providing comments asking for more from these events were specifically geared toward having the events closer to where they reside, indicating that there is a call and need for these events.

Promoting the events on Facebook also allowed for conversations between users and Seattle Children's Hospital, the host of the events. Facebook users were able to comment with questions about the event and either other users or the host could respond in a more immediate manner. Findings from this study also support the presence of certain non-host Facebook users or "moderators," who speak up to draw the conversations back to the goal of the event (ie, to protect children and households through safe firearm storage practices). Moreover, these individuals facilitate the conversation in a direction that deters the discussion away from topics such as the legalities of firearm ownership.

Although there were some negative and advocacy-oriented comments, there was an emergence of users who implicitly function as moderators on Facebook along with ability for the host to respond and moderate as well. Moderators have the potential to also lift some of the burdens off the event host to have to constantly monitor and intervene. Previous studies have found that moderators in online communities have a substantial contribution in keeping discussions on track by countering negative social media perceptions [25]. This observation begins to highlight the emerging importance of engaging positive influencers on social media for promotion of events such as the SFSGOE. Other studies have also highlighted the importance of having "moderated" discussions, which have been shown to yield better participant engagement and to promote a safer discussion environment [26,27].

We also observed a much higher level of positive feedback than negative feedback. There were more interactions such as clicking and liking the post compared to hiding or unliking the post. This high level of positive responsiveness is important because it shows that despite the controversial subject matter, no notable backlash toward the organization hosting the event or any perceivable impact on their clinically oriented branding occurred.

Over one-third of all participants across the 13 events found out about the event through social media, indicating that this is a useful marketing tool. Additionally, more of the younger participants and those who identified as female found out about these events through social media than through the other promotion tools such as flyers, newspapers, and word of mouth. Those who found out about the event on social media were on average 6.4 years younger than those who found out through other means. Contextually, this represents the users of Facebook, in which there is a higher proportion of female users compared to male users (74% vs 62% of US adults) and a decreasing trend in usage with increasing age [28]. This is important when considering the intended demographic characteristics of reach.

There was no significant difference in current firearm storage practices between the social media group and the other group. Participants who found out about the event on social media expressed an even greater intention to use the giveaway event storage device than those who found out about the event through other means. When implementing an intervention, it is important to use the most appropriate and efficient tools to promote it. This study suggests that social media captures younger individuals and a different gender distribution than means such as flyers, newspaper ads, and posters. This finding further supports other social media and health promotion studies that have also found social media to be an acceptable communication tool; however, the effectiveness varies across different demographic groups [29,30].

Using a concurrent mixed methods approach, the integration of the qualitative and quantitative methodology allowed for results from both arms of the study to complement and support each other. The social media analysis allowed for gaining an understanding of receptiveness and responsiveness, and we were further able to use the data from the survey to understand the type of people that social media promotions reached and the characteristics of that population.

However, due to the limited qualitative and quantitative data for rural locations, we were unable to determine whether social media would be able to improve the reach in marketing safe firearm storage promotion events.

Limitations

In using Facebook data to understand receptiveness, there are likely to be biases in the attitudes presented. Across social media platforms, including Facebook, there is a tendency for greater positive expressions because they are generally perceived as more appropriate than negative sentiments [31]. Similarly, we found many more positive comments than negative comments. Facebook users with negative perceptions toward firearm storage may not find Facebook to be the most appropriate platform to express that sentiment, since it is a publicly facing platform and also amidst comments that lean more heavily toward the positive. Additionally, there is the chance that certain comments were deleted or reported by Facebook users by the time we were able to retrospectively export data for this analysis. To our knowledge, no comments were removed by Seattle Children's Hospital.

Due to the finite number of comments publicly available for this exploratory content analysis, we have limited understanding of the saturation of these codes. With the inductive thematic saturation approach, we were able to generate seven themes; however, future studies with greater elements of coding would allow gaining a better understanding of the reproducibility of use in this exploratory inductively generated codebook [22].

The social media content analysis evaluated data from 13 events, although 15 events occurred between December 2014 and March 2018. There were two events that were not evaluated because they were not publicly available on Facebook at the time of data extraction. This decreased the sample size for our qualitative analysis and also may have decreased the generalizability of the results due to the homogeneity of the settings. Of the 13 events evaluated qualitatively, 8 were held in larger cities from urban counties and only 1 was located in a city from a semirural county. Additionally, through the quantitative portions of the social media evaluation, an additional 4 events did not yield data through SimplyMeasured, the social media analytics software. Because of these missing data, there may have been a loss in positive feedback data as well as negative feedback data. Given this level of missing social media data, the findings described in this study may only be generalizable to events held in urban areas.

Additionally, of the 5512 records of surveys collected, there was an 81.7% (4504/5512) completion response rate. We defined a complete survey as one with location of the event attended indicated and that was marked as "complete" by the data entry program staff. There could be a systematic difference between those who completed the survey and those who chose not to complete any or certain fields of the survey. Moreover, with the analyzed data, we found much higher levels of missingness among respondents who indicated that they found out about the events through means other than social media. Dependent on the characteristic of interest, the proportion of missing data ranged from 4.9% to 12.2%. An added limitation of the results from the survey analysis is that, due to its anonymous nature, it is possible that participants could have attended more than one event; therefore, their characteristics would have been counted more than once during aggregate analysis. We did not have identifiers to disaggregate and remove any repeated responses from the same participant.

For the purposes of this study, we used the location of the event as a proxy to generally understand if the participants were from rural or urban areas. However, we were unable to truly ascertain the city or county that the participants were traveling from to attend these events. This was because location of residence was not asked in the survey for the earlier events, and there were very high levels of missing data from the surveys completed in the latter events. Obtaining information about where participants are coming from would have better informed not only how far the participants were traveling but also how social media marketing may be reaching communities closer to or further from the location of these events.

Conclusions

We found high levels of positive receptiveness and responsiveness toward event posts promoting SFSGOE. Specifically, these social media posts reached a greater proportion of younger participants. Among participants who identified as female, a greater proportion found out about these events through social media. Due to missing data and stronger data skew for urban area events, we were unable to draw conclusions as to whether or not there were differences in acceptability of such events or using social media to market these events between rural and urban areas.

Social media allowed for multifaceted interactions from user to user, from the user to the host, and from the user to the social media platform. We also found the emergence of Facebook "moderators" whose implicit role helped in supporting the goals of the health promotion event. These interactions are unique because social media facilitate more immediate displays of receptiveness and responsiveness compared with more traditional methods of health promotion, such as posters and flyers. Our analysis showed high levels of acceptability in using social media to market an intervention that may have more controversial connotations. Additionally, our findings support the idea that social media can be used as a tool for health promotion, specifically as a tool for promoting and discussing community-focused interventions. Future research should focus on understanding if there is a difference in rural vs urban acceptability for firearm storage education and promotion, and on the greater implementation of these events to encourage safe firearm storage practices. Additionally, further research can expand the understanding of implementing health promotion activities on social media platforms beyond Facebook, such as on Twitter and Instagram.

Acknowledgments

The authors would like to thank Seattle Children's Hospital Community Health and Benefit and Marketing and Communications teams for their support in this work, specifically Stacy Andrews, Cassie Simeona, Isabell Sakamoto, and Nisreen El-Saadoun for their help in data management and manuscript review. Additionally, we would like to thank Aubrey Gower from the University of Wisconsin-Madison for support with the qualitative data analysis portions of this project. Funding for the SFSGOE was provided by Seattle Children’s Hospital.

Conflicts of Interest

None declared.

Multimedia Appendix 1
SFSGOE Intake Survey.
[PDF File (Adobe PDF File), 307 KB - jmir_v23i6e24458_app1.pdf]

References

1. Fowler KA, Dahlberg LL, Haileyesus T, Annest JL. Firearm injuries in the United States. Prev Med 2015 Oct;79:5-14 [FREE Full text] [doi: 10.1016/j.ypmed.2015.06.002] [Medline: 26116133]
2. Fowler KA, Dahlberg LL, Haileyesus T, Gutierrez C, Bacon S. Childhood firearm injuries in the United States. Pediatrics 2017 Jul;140(1):e20163486 [FREE Full text] [doi: 10.1542/peds.2016-3486] [Medline: 28630118]
3. Fontanella CA, Hiance-Steelesmith DL, Phillips GS, Bridge JA, Lester N, Sweeney HA, et al. Widening rural-urban disparities in youth suicides, United States, 1996-2010. JAMA Pediatr 2015 May;169(5):466-473 [FREE Full text] [doi: 10.1001/jamapediatrics.2014.3561] [Medline: 25751611]
4. Herrin BR, Gaither JR, Leventhal JM, Dodington J. Rural versus urban hospitalizations for firearm injuries in children and adolescents. Pediatrics 2018 Aug;142(2):e20173318 [FREE Full text] [doi: 10.1542/peds.2017-3318] [Medline: 29967056]
5. 2018 Washington State Health Assessment - Suicide & Safe Storage of Firearms. URL: https://www.doh.wa.gov/Portals/1/Documents/Pubs/78945-SHA.pdf [accessed 2019-05-19]
6. Grossman DC, Mueller BA, Riedy C, Dowd MD, Villaveces A, Prodzinski J, et al. Gun storage practices and risk of youth suicide and unintentional firearm injuries. JAMA 2005 Feb 09;293(6):707-714. [doi: 10.1001/jama.293.6.707] [Medline: 15701912]
7. Coker AL, Bush HM, Follingstad DR, Brancato CJ. Frequency of guns in the households of high school seniors. J Sch Health 2017 Mar;87(3):153-158 [FREE Full text] [doi: 10.1111/josh.12479] [Medline: 28147456]
8. Simonetti JA, Mackelprang JL, Rowhani-Rahbar A, Zatzick D, Rivara FP. Psychiatric comorbidity, suicidality, and in-home firearm access among a nationally representative sample of adolescents. JAMA Psychiatry 2015 Feb;72(2):152-159. [doi: 10.1001/jamapsychiatry.2014.1760] [Medline: 25548879]
9. Azrael D, Cohen J, Salhi C, Miller M. Firearm storage in gun-owning households with children: results of a 2015 National Survey. J Urban Health 2018 Jun;95(3):295-304 [FREE Full text] [doi: 10.1007/s11524-018-0261-7] [Medline: 29748766]
10. Morgan ER, Gomez A, Rowhani-Rahbar A. Firearm ownership, storage practices, and suicide risk factors in Washington state, 2013-2016. Am J Public Health 2018 Jul;108(7):882-888. [doi: 10.2105/AJPH.2018.304403] [Medline: 29771611]
11. King A, Simonetti J, Bennett E, Simeona C, Stanek L, Roxby AC, et al. Firearm storage practices in households with children: A survey of community-based firearm safety event participants. Prev Med 2020 Feb;131:105952. [doi: 10.1016/j.ypmed.2019.105952] [Medline: 31843466]
12. Washington State Initiative Measure No. 1639. Secretary of the State. URL: https://www.sos.wa.gov/_assets/elections/initiatives/finaltext_1531.pdf [accessed 2018-12-19]
13. Simonetti JA, Rowhani-Rahbar A, King C, Bennett E, Rivara FP. Evaluation of a community-based safe firearm and ammunition storage intervention. Inj Prev 2018 Jun;24(3):218-223. [doi: 10.1136/injuryprev-2016-042292] [Medline: 28642248]
14. Smith WA. Social marketing: an overview of approach and effects. Inj Prev 2006 Jun;12(Suppl 1):i38-i43 [FREE Full text] [doi: 10.1136/ip.2006.012864] [Medline: 16788110]
15. Lee N, Kotler P. Social Marketing: Influencing Behaviors for Good. 4th Edition. Newbury Park, CA: SAGE Publications; 2011.
16. Albright K, Gechter K, Kempe A. Importance of mixed methods in pragmatic trials and dissemination and implementation research. Acad Pediatr 2013;13(5):400-407. [doi: 10.1016/j.acap.2013.06.010] [Medline: 24011744]
17. Andrade EL, Evans WD, Barrett N, Edberg MC, Cleary SD. Strategies to increase Latino immigrant youth engagement in health promotion using social media: mixed-methods study. JMIR Public Health Surveill 2018 Dec 19;4(4):e71 [FREE Full text] [doi: 10.2196/publichealth.9332] [Medline: 30567689]
18. Partridge SR, Grunseit AC, Gallagher P, Freeman B, O'Hara BJ, Neubeck L, et al. Cardiac patients' experiences and perceptions of social media: mixed-methods study. J Med Internet Res 2017 Sep 15;19(9):e323 [FREE Full text] [doi: 10.2196/jmir.8081] [Medline: 28916507]
19. 2013 Rural-Urban Continuum Codes. Economic Research Service, United States Department of Agriculture. 2013. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx [accessed 2018-01-20]
20. Neale J. Iterative categorization (IC): a systematic technique for analysing qualitative data. Addiction 2016 Jun;111(6):1096-1106 [FREE Full text] [doi: 10.1111/add.13314] [Medline: 26806155]

21. Thomas D. A general inductive approach for qualitative data analysis. Am J Eval 2003;27:237-246.
22. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
24. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
25. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017 Oct;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
26. Windler C, Clair M, Long C, Boyle L, Radovic A. Role of moderators on engagement of adolescents with Depression or anxiety in a social media intervention: content analysis of web-based interactions. JMIR Ment Health 2019 Sep 26;6(9):e13467 [FREE Full text] [doi: 10.2196/13467] [Medline: 31573923]
27. Kendal S, Kirk S, Elvey R, Catchpole R, Pryjmachuk S. How a moderated online discussion forum facilitates support for young people with eating disorders. Health Expect 2017 Feb;20(1):98-111 [FREE Full text] [doi: 10.1111/hex.12439] [Medline: 26725547]
28. Pew Research Center - Internet and Technology. Social Media Fact Sheet. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2020-05-20]
29. Klassen KM, Douglass CH, Brennan L, Truby H, Lim MSC. Social media use for nutrition outcomes in young adults: a mixed-methods systematic review. Int J Behav Nutr Phys Act 2018 Jul 24;15(1):70 [FREE Full text] [doi: 10.1186/s12966-018-0696-y] [Medline: 30041699]
30. Welch V, Petkovic J, Pardo Pardo J, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75. [doi: 10.24095/hpcdp.36.4.01] [Medline: 27077792]
31. Waterloo SF, Baumgartner SE, Peter J, Valkenburg PM. Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. New Media Soc 2018 May;20(5):1813-1831 [FREE Full text] [doi: 10.1177/1461444817707349] [Medline: 30581358]

Abbreviations

RUCC: Rural-Urban Continuum Code
SFSGOE: Safe Firearm Storage Giveaway Outreach Event

Edited by R Kukafka; submitted 25.09.20; peer-reviewed by K Carlson, T Alessa; comments to author 26.01.21; revised version received 19.03.21; accepted 24.03.21; published 18.06.21.

Please cite as:
Lam E, Moreno M, Bennett E, Rowhani-Rahbar A
Receptiveness and Responsiveness Toward Using Social Media for Safe Firearm Storage Outreach: Mixed Methods Study
J Med Internet Res 2021;23(6):e24458
URL: https://www.jmir.org/2021/6/e24458
doi:10.2196/24458
PMID:34142974

Original Paper

Validation of Visual and Auditory Digital Markers of Suicidality in Acutely Suicidal Psychiatric Inpatients: Proof-of-Concept Study

Isaac Galatzer-Levy[1,2], PhD; Anzar Abbas[1], PhD; Anja Ries[3,4,5], PhD; Stephanie Homan[3,4,5], PhD; Laura Sels[3,4], PhD; Vidya Koesmahargyo[1], MSc; Vijay Yadav[1], MSc; Michael Colla[4], MD; Hanne Scheerer[4], MD; Stefan Vetter[4], MD; Erich Seifritz[4,5], MD; Urte Scholz[4], PhD; Birgit Kleim[3,4,5]

[1]Research and Development, AiCure, New York, NY, United States
[2]Psychiatry, New York University School of Medicine, New York, NY, United States
[3]Department of Psychology, University of Zurich, Zurich, Switzerland
[4]Department of Psychiatry, Psychotherapy and Psychosomatics, University of Zurich, Zurich, Switzerland
[5]Neuroscience Centre Zurich, University of Zurich, Zurich, Switzerland

Corresponding Author:
Vidya Koesmahargyo, MSc
Research and Development
AiCure
19 W 24th St
11th Floor
New York, NY, 10010
United States
Phone: 1 6463015037
Email: vidya.koesmahargyo@aicure.com

Abstract

Background: Multiple symptoms of suicide risk have been assessed based on visual and auditory information, including flattened affect, reduced movement, and slowed speech. Objective quantification of such symptomatology from novel data sources can increase the sensitivity, scalability, and timeliness of suicide risk assessment.

Objective: We aimed to examine measurements extracted from video interviews using open-source deep learning algorithms to quantify facial, vocal, and movement behaviors in relation to suicide risk severity in recently admitted patients following a suicide attempt.

Methods: We utilized video to quantify facial, vocal, and movement markers associated with mood, emotion, and motor functioning from a structured clinical conversation in 20 patients admitted to a psychiatric hospital following a suicide risk attempt. Measures were calculated using open-source deep learning algorithms for processing facial expressivity, head movement, and vocal characteristics. Derived digital measures of flattened affect, reduced movement, and slowed speech were compared to suicide risk with the Beck Scale for Suicide Ideation controlling for age and sex, using multiple linear regression.

Results: Suicide severity was associated with multiple visual and auditory markers, including *speech prevalence* ($\beta=-0.68$, $P=.02$, $r^2=0.40$), *overall expressivity* ($\beta=-0.46$, $P=.10$, $r^2=0.27$), and head movement measured as *head pitch variability* ($\beta=-1.24$, $P=.006$, $r^2=0.48$) and *head yaw variability* ($\beta=-0.54$, $P=.06$, $r^2=0.32$).

Conclusions: Digital measurements of facial affect, movement, and speech prevalence demonstrated strong effect sizes and linear associations with the severity of suicidal ideation.

(J Med Internet Res 2021;23(6):e25199) doi:10.2196/25199

KEYWORDS

digital phenotyping; digital biomarkers; digital health; depression; suicidal ideation; digital markers; digital; facial; suicide; suicide risk; visual; auditory

Introduction

Timely, efficient, sensitive, and noninvasive measurement approaches are required to improve suicide risk assessment [1]. One promising avenue is the use of remote digital monitoring methodologies that utilize smart devices, such as phones and wearables, in conjunction with deep learning algorithms to infer behavioral and physiological states associated with suicide risk [2,3].

While suicide risk is often comorbid with other mental disorders, in particular, major depressive disorder (MDD) and schizophrenia, suicidal behavior is increasingly recognized as unique in presentation and risk profile [4-7]. Based on prior knowledge, visual and auditory data sources represent a compelling direction in the objective measurement of behavior associated with suicide risk. Emil Kraepelin first observed that suicide risk was associated with melancholic states characterized by slowed speech, where patients appeared to “become mute in the middle of a sentence,” and further observed how “the facial expression and the general attitude are sleepy and languid, the speech is low…” [8]. More contemporarily, reduced facial expressivity and movement measured using standardized coding schemes based on videos of patient interviews differentiated depressed patients with and without suicide risk [9], and altered vocal characteristics have been observed in acutely suicidal patients [5].

A number of visual and auditory characteristics can be directly quantified, including gross motor activity [10], head movement variability [11-13], facial activity [14], and properties of speech [15]. The automated measurement of these clinical features introduces the possibility of objective digital assessment of visual and auditory markers of suicide risk. Given that audio and video data sources can be captured remotely, this further introduces the possibility of greatly scaling the reach and frequency of assessment. Increased scale and objectivity can facilitate increased accuracy and accessibility of clinical risk and treatment response assessment.

Visual and auditory biomarkers were selected based on a mechanistic theory that reduced serotonin, a key risk factor for suicidal behavior and a primary biological target for treatment of MDD, will affect behavioral characteristics, including an individual’s speech, head movement, and facial activity. Serotonergic tone is known to mechanistically impact motor functioning directly and via interactions with dopamine and norepinephrine signaling [16-18]. Depleted serotonin has been observed in the postmortem brains of individuals with MDD [19] and those who have committed suicide [20]. Furthermore, direct pharmacological manipulation of serotonin using selective serotonin reuptake inhibitors (SSRIs) increases suicide risk [21,22].

While novel measurements are promising, validation is required before such metrics can be interpreted clinically. Key steps for validation include comparison to traditional clinical measures, both cross-sectionally and as they change alongside treatment and disease course [23]. Such measures should strive to be easy to collect, should have increased sensitivity to facilitate frequent and accurate assessment, and should be validated in relationship to narrower biological phenotypes and treatment targets than traditional endpoints offer. This will lead to improved dynamic treatment research and clinical decision making based on modulation of underlying neurobiological deficits [24,25].

In this study, we examined measurements extracted from video interviews using open-source deep learning algorithms to quantify facial, vocal, and movement behaviors in relation to suicide risk severity in patients interviewed soon after admission to a psychiatric hospital following a suicide attempt.

Methods

Study Participants

Participants were recruited from the Psychiatric University Hospital, Zurich, Switzerland, as part of the “SIMON–Suicide Ideation MONitoring” study. The SIMON study has been funded by the Digital Lives initiative of the Swiss National Science Foundation (SNSF) and was approved by the ethics committee of the Faculty of Philosophy, University of Zurich (approval number: 19.2.1). The study aims to develop a digital protocol to monitor and predict suicidal ideation and hospital readmission in high-risk psychiatric patients. Digital markers of psychopathology are set to be derived from smartphone-based experience sampling, mobile passive sensing, and video recordings of patient free speech.

Patients were included in the study if they met the following criteria: (1) admission to the hospital after a suicide attempt or in the context of suicidal ideation, and suicidal ideation was identified in the first diagnostic intake interview, (2) sufficient knowledge of the German language, (3) being a smartphone user, and (4) discharge from the clinic after identification of suicidal ideation, with established outpatient care contact with a physician or psychologist. Patients were excluded if they met the following criteria: (1) planning to leave the greater Zurich area within the study period, (2) sharing a smartphone with another person, and (3) being active military personnel. Researchers kept track of all patients admitted to the hospital and contacted the treating psychologist or physician in case of eligibility. Patients who met the inclusion criteria and for whom an approval from the psychologist or physician was granted were informed about the study. If patients agreed to participate in the study, they were invited for a baseline assessment appointment that entailed the following: (1) detailed information about the study, (2) collection of informed consent signed by the patient, (3) assessment of current mental disorders with the Mini International Neuropsychiatric Interview (MINI; version 6), (4) a short videotaped semistructured qualitative interview (for which additional informed consent signed by the patient was obtained) upon agreement, (5) electronic/pen and paper questionnaires evaluating relevant psychological variables, and (6) smartphone app installation. Participants were reimbursed with up to CHF 120 (US $127).

At the time of the video assessment, all patients had an inpatient status at the Psychiatric University Hospital Zurich. We recruited patients admitted to the psychiatric hospital to ensure an appropriate reach and a sufficient sample size of the high-risk psychiatric patient group. For practical reasons and because of the format of the semistructured interviews (ie, questions asked by the researcher and video recordings with a tablet), we obtained video recordings during the baseline interview, when patients were still inpatients. However, the remaining parts of the study (ie, remote patient monitoring using smartphones) commenced after patients were discharged from the hospital. In that way, we assessed patient well-being and behavior in an ecologically valid manner outside the hospital. However, this study only focused on the video recordings from the time patients were inpatients.

Participants were recruited for the study on a rolling basis. At the time of the analysis, 30 patients agreed to participate in the videotaped interview, of whom 20 completed the necessary baseline questionnaires. Ultimately, the analysis was conducted on a sample of 20 patients.

Clinical Assessment of Suicidality

Assessment of suicide risk was completed at baseline. The Beck Scale for Suicide Ideation (BSS) questionnaire (German validated version) [26] was administered to evaluate patients' current intensity of attitudes, plans, and behaviors to commit suicide. Patients' history of nonsuicidal self-injury and suicide attempts was assessed using the following two self-report items from the Self-Injurious Thoughts and Behaviors Interview (SITBI): (1) "In your life, have you purposefully hurt yourself without wanting to die?" and (2) "How many times in your lifetime have you made an attempt to kill yourself during which you had at least some intent to die?" [27,28].

Collection of Video and Audio Data

At baseline, patients were given the choice of additionally participating in a video-recorded interview. Upon agreement, patients signed an informed consent form specific to this part of the study. A short videotaped semistructured qualitative interview was performed. A laptop was placed on the table in front of the patient, and 1-minute video and audio samples were recorded. During the qualitative video interview, participants answered introductory questions assessing their current state, as well as questions about experiences with different valences (ie, neutral, positive, and negative) and temporal dimensions (ie, past, present, and future). Overall, the following six video and audio samples, each at 1 minute, were recorded per participant: introduction (neutral present), neutral, positive past, positive future, negative past, and negative future. Multimedia Appendix 1 displays exemplary questions asked for each category. The conversation for each category starts with the central question (first in order) asked by the researcher, followed by additional questions to keep the participant involved and talking during the 1-minute recordings.

Machine Learning Computer Vision and Voice Analysis

All analyses were conducted in a Python environment using open-source tools. No novel machine learning models were trained for the purposes of this study, rather existing tools for the measurement of facial expressivity [29], vocal acoustics [30], and patterns of movement [31] were utilized. The code used to calculate the biomarkers has been compiled into its own open-source package, which allows for other researchers to replicate the methods used in this manuscript [32]. The videos were segmented to only include the participants' responses to the questionnaires, with behaviors during free speech cropped out and then concatenated for analysis of digital markers.

Facial Activity Analysis

Initially, all videos were segmented into frames, resulting in a minimum of 33 image frames per second. Thereafter, OpenCV, an open-source computer vision software package [33], was used to segment each image into three matrices consisting of red, blue, and green spectrum pixels. Subsequently, each frame was analyzed using OpenFace [29], an open source software package that has demonstrated validity next to expert human ratings of Facial Action Coding Scheme (FACS) [34], a standardized methodology to measure facial movements that reflect activity in the underlying human facial musculature utilized in the production of basic emotions. Specifically, for each frame, OpenFace outputs (1) a binary value indicating the presence of an action unit and (2) a continuous value indicating the intensity with which the action unit is being expressed. Action unit intensities were used to derive measures of individual emotional expressivity (ie, *happiness expressivity*, *sadness expressivity*, *fear expressivity*, *anger expressivity*, *surprise expressivity*, and *disgust expressivity*). Action unit intensities were also used to calculate *overall expressivity* regardless of emotion.

Movement Analysis

The angle of the head's pitch (up and down movement) and yaw (side to side movement) was acquired for each video frame using OpenFace. The standard deviations of the frame-wise pitch and yaw measurements were used as indicators of head movement (ie, *head pitch variability* and *head yaw variability*).

Voice Analysis

Recordings were segmented into speech and nonspeech parts using Parselmouth, an open-source software package utilized for vocal analysis [30]. The percentage of the audio file where speech was recorded as opposed to no speech (ie, *speech prevalence*) was calculated to represent how much the participant was talking given the length of the recording.

Data Analysis

Initially, BSS scores were regressed on all movement, facial, and audio markers controlling for age and sex using separate multiple linear regressions. Facial activity, movement, and voice are all behavioral characteristics that are influenced by both age and sex. Hence, it was important to conduct a multiple linear regression controlling for age and sex in order to remove the influence of those factors on the final comparisons between digital measures and BSS scores. In addition to significance testing, unique variance accounted for in BSS scores by the digital measures was assessed using Cohen *d* [35]. Following analysis of *overall expressivity*, the BSS was regressed on expressivity of each emotion correcting for multiple comparisons using false discovery rate adjustment to reduce *P*-value inflation due to chance.

Results

Suicide Severity

Suicidal ideation severity on the BSS across subjects was, on average, above the clinical cutoff of 9, indicating severe suicide risk (μ=10.9, σ=10.2, range 0-31) and an average count of three past suicide attempts (σ=6.9, range 0-30).

Facial Activity

Controlling for age and sex, *overall expressivity* demonstrated a significant negative linear association with BSS scores, indicating that decreased facial activity is associated with greater suicide risk (β=−0.46, *P*=.01, r^2=0.27). This indicates that facial activity decreases as suicide severity increases, with the strongest evidence in the context of facial expressions without emotional expressions.

To better understand if particular emotions contributed to the observation of *composite expressivity*, we regressed BSS scores with individual emotional expressivity, sex, and age. *P* values were adjusted using false discovery rate adjustment. Post-hoc analyses demonstrated significance using Benjamini-Hochberg adjusted *P* values for *sadness expressivity* (β=−0.68, *P*=.01, r^2=.43), *surprise expressivity* (β=−0.74, *P*=.002, r^2=0.53), and *disgust expressivity* (β=−0.64, *P*=.04, r^2=0.35), but not for *fear*, *anger*, and *happiness expressivity* (Table 1).

Table 1. Results from multiple linear regression between Beck Scale for Suicide Ideation questionnaire scores and digital measurements of facial expressivity, vocal behavior, and head movement, controlling for age and sex.

Predictor	Coefficient	Standard error	*t* (df)	*P* value	*F* (df)	R^2	Adjusted R^2	*P* value
Regression 1					0.9013 (1)	0.153	−0.017	.46
Constant	0.2020	0.380	0.531 (18)	.60				
Anger expressivity	−0.2456	0.322	−0.762 (18)	.46				
Age	0.0068	0.007	0.960 (18)	.35				
Sex	0.1167	0.172	0.680 (18)	.51				
Regression 2					2.6930 (1)	0.350	0.220	.08
Constant	0.1213	0.240	0.506 (18)	.62				
Disgust expressivity	−0.6401	0.278	−2.305 (18)	.04				
Age	0.0120	0.006	1.980 (18)	.07				
Sex	0.1422	0.146	0.972 (18)	.35				
Regression 3					1.0430 (1)	0.173	0.007	.40
Constant	0.2628	0.380	0.692 (18)	.50				
Fear expressivity	−0.3210	0.328	−0.978 (18)	.34				
Age	0.0060	0.007	0.842 (18)	.41				
Sex	0.1056	0.170	0.620 (18)	.54				
Regression 4					0.6834 (1)	0.120	−0.056	.58
Constant	0.0051	0.292	0.018 (18)	.99				
Happiness expressivity	−0.0228	0.255	−0.089 (18)	.93				
Age	0.0083	0.007	1.194 (18)	.25				
Sex	0.1517	0.178	0.852 (18)	.41				
Regression 5					3.7140 (1)	0.426	0.311	.04
Constant	0.4394	0.270	1.629 (18)	.12				
Sadness expressivity	−0.6785	0.240	−2.830 (18)	.01				
Age	0.0082	0.006	1.480 (18)	.16				
Sex	0.0345	0.152	−0.228 (18)	.82				
Regression 6					5.7130 (1)	0.533	0.440	.008
Constant	−0.0441	0.198	−0.223 (18)	.83				
Surprise expressivity	−0.7437	0.204	−3.645 (18)	.002				
Age	0.0143	0.005	2.734 (18)	.02				
Sex	0.2421	0.127	1.912 (18)	.08				
Regression 7					1.8200 (1)	0.267	0.120	.19
Constant	0.2881	0.300	0.961 (18)	.35				
Overall expressivity	−0.4585	0.264	−1.734 (18)	.10				
Age	0.0054	0.006	0.833 (18)	.42				
Sex	0.1947	0.158	1.234 (18)	.24				
Regression 8					3.3890 (1)	0.404	0.285	.046
Constant	0.3291	0.256	1.285 (18)	.22				
Speech prevalence	−0.6808	0.255	−2.674 (18)	.02				
Age	0.0057	0.006	0.991 (18)	.34				
Sex	0.3429	0.158	2.170 (18)	.047				
Regression 9					4.5330 (1)	0.476	0.371	.02

Predictor	Coefficient	Standard error	*t* (df)	*P* value	*F* (df)	R^2	Adjusted R^2	*P* value
Constant	0.4192	0.248	1.689 (18)	.11				
Head pitch variability	−1.2465	0.391	−3.189 (18)	.006				
Age	0.0124	0.005	2.294 (18)	.04				
Sex	−0.0342	0.143	−0.239 (18)	.81				
Regression 10					0.1170 (1)	0.317	0.181	.12
Constant	0.1006	0.245	0.411 (18)	.69				
Head yaw variability	−0.5408	0.260	−2.081 (18)	.06				
Age	0.0125	0.006	1.979 (18)	.07				
Sex	0.1723	0.150	1.146 (18)	.27				

Voice Analysis

Controlling for age and sex, *speech prevalence* demonstrated a significant negative linear association with BSS scores, indicating that suicide severity scores increased as speech decreased ($\beta=-0.68$; $P=.02$, $r^2=0.40$; Table 1).

Movement Analysis

Controlling for age and sex, both *head pitch variability* and *head yaw variability* demonstrated significant negative linear relationships with suicide severity ($\beta_{pitch}=-1.24$, $P=.006$, $r^2=0.48$; $\beta_{yaw}=-0.54$, $P=.055$, $r^2=0.32$), indicating that lower levels of head movement were associated with greater suicide severity scores (Table 1).

Discussion

We examined visual and auditory measures of facial activity, head movement, and speech production, which were all calculated using deep learning algorithms applied to open-ended clinical interviews with psychiatric patients following a suicide attempt. The goal of this work was to determine if key indicators of suicide severity could be measured in an objective and automated manner using video data captured during clinical interviews that provided structured questions, but were otherwise kept deliberately open to mimic psychiatric interviewing in routine care. Achieving this goal provides a proof of concept that suicide risk can be assessed through analysis of unstructured video interview data in conjunction with deep learning algorithms designed to measure clinical/behavioral characteristics.

Objective measurements of visual and auditory markers of suicidality calculated from videos of patient behavior can be useful in the context of psychiatric care and clinical research, in particular, if the measurements can be acquired without specialized hardware using video or audio from regular patient-clinician interactions, as has been demonstrated in this study. Such tools have applications in clinical practice as they can allow for remote measurement of symptomatology. For example, virtual clinic visits through video calls can involve digital measurements to assess mental health if such measurements are integrated into the technology utilized. Existing smartphone-based platforms allow for the collection of video and audio data and subsequent calculation of digital measures, and can be applied in out-patient settings to remotely acquire these measurements without adding further burden on the clinical staff [36-38]. Of course, such technologies face a great deal of challenges before they can be deployed in patient care, starting with further validation of methodologies, passage through regulatory review as dependable clinical measures, and adoption by health care professionals into their day-to-day clinical workflows [39]. Ultimately, methods that can provide remote, objective, and passive measurement of clinical functioning can greatly increase the reach of treatment development, dissemination, and personalization [25,40].

Finally, the studied markers were selected based on a mechanistic theory that reduced serotonin in the brain, a key risk factor for suicidal behavior, will manifest in behavioral characteristics, including those involving a patient's speech, head movement, and facial activity. These primary hypotheses were confirmed. Overall facial expressivity, speech activity, and head movement during spontaneous behavior, which are all downstream effects of motor slowing, were correlated with BSS scores. This work provides a proof of concept that proxy markers of underlying neurobiological functioning can be used to measure clinical risk. Demonstrating this indicates that there is potential to remotely measure more specific neurobiological phenotypes to determine risk or examine the response to treatment. For example, owing to suicide risk associated with SSRIs, there is a need for close clinical monitoring during the initiation of treatment. Similarly, such models may be useful to determine who would benefit from treatments that affect serotonergic tone. Ultimately, by measuring a more specific biologically based phenotype, there is greatly increased opportunity to improve sensitivity of measurement and specificity of treatment [24].

The work has key limitations. First, this work represents a proof of concept. The indicated markers for risk assessment or measurement of clinical severity should only be used in an exploratory manner until such markers can be validated in larger and more diverse clinical populations and settings. We need to directly determine the amount of video data needed and the clinically meaningful cutoffs to determine clinical functioning. It is likely that like traditional suicide risk assessment, multiple characteristics together are needed to determine risk or severity. Multiple digital measures can be combined to produce a more robust metric, but like traditional scale development, this requires larger samples from diverse populations. Importantly, the current work uses open-source software, and additionally, we have published our software methods. As such, there is no access barrier for researchers to implement identical methods to test and extend the current approach. An open-source approach lends itself to rapid replication, extension, and implementation.

Ultimately, this proof-of-concept study demonstrates that theoretically grounded visual and auditory digital measurements are valid as markers of suicide severity. This effort provided a digital approach akin to traditional clinical assessment whereby a skilled clinician listens to a patient and applies internal working models developed through years of experience to assess clinical functioning.

Conflicts of Interest

AA, VK, and VY are employed by and hold stock options in AiCure, LLC. The other authors have no conflicts to declare.

Multimedia Appendix 1

Exemplary questions for the six categories of the videotaped semistructured qualitative interview.

[DOCX File , 21 KB - jmir_v23i6e25199_app1.docx]

References

1. Claassen CA, National Action Alliance for Suicide Prevention Research Prioritization Task. The agenda development process of the United States' National Action Alliance for Suicide Prevention Research Prioritization Task Force. Crisis 2013 May;34(3):147-155. [doi: 10.1027/0227-5910/a000208] [Medline: 23628664]
2. Torous J, Walker R. Leveraging Digital Health and Machine Learning Toward Reducing Suicide-From Panacea to Practical Tool. JAMA Psychiatry 2019 Oct 01;76(10):999-1000 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.1231] [Medline: 31290952]
3. Vahabzadeh A, Sahin N, Kalali A. Digital Suicide Prevention: Can Technology Become a Game-changer? Innov Clin Neurosci 2016;13(5-6):16-20 [FREE Full text] [Medline: 27800282]
4. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Washington, DC: American Psychiatric Association; 2013.
5. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Communication 2015 Jul;71:10-49. [doi: 10.1016/j.specom.2015.03.004]
6. Pompili M, Amador XF, Girardi P, Harkavy-Friedman J, Harrow M, Kaplan K, et al. Suicide risk in schizophrenia: learning from the past to change the future. Ann Gen Psychiatry 2007 Mar 16;6(1):1-22. [doi: 10.1186/1744-859x-6-10]
7. Sisti D, Mann JJ, Oquendo MA. Toward a Distinct Mental Disorder-Suicidal Behavior. JAMA Psychiatry 2020 Jul 01;77(7):661-662. [doi: 10.1001/jamapsychiatry.2020.0111] [Medline: 32186693]
8. Diefendorf AR, Kraepelin E. Clinical psychiatry: A textbook for students and physicians, abstracted and adapted from the 7th German edition of Kraepelin's Lehrbuch der Psychiatrie. New York, NY: MacMillan Co; 1907.
9. Heller M, Haynal V. Depression and Suicide Faces. In: Ekman P, Rosenberg EL, editors. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford, United Kingdom: Oxford University Press; 2005.
10. Buyukdura JS, McClintock SM, Croarkin PE. Psychomotor retardation in depression: biological underpinnings, measurement, and treatment. Prog Neuropsychopharmacol Biol Psychiatry 2011 Mar 30;35(2):395-409 [FREE Full text] [doi: 10.1016/j.pnpbp.2010.10.019] [Medline: 21044654]
11. Alghowinem S, Goecke R, Wagner M, Parkerx G, Breakspear M. Head Pose and Movement Analysis as an Indicator of Depression. 2013 Presented at: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction; September 2-5, 2013; Geneva, Switzerland. [doi: 10.1109/acii.2013.53]
12. Dibeklioglu H, Hammal Z, Cohn JF. Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding. IEEE J. Biomed. Health Inform 2018 Mar;22(2):525-536. [doi: 10.1109/jbhi.2017.2676878]
13. Kim Y, Cheon S, Youm C, Son M, Kim JW. Depression and posture in patients with Parkinson's disease. Gait Posture 2018 Mar;61:81-85. [doi: 10.1016/j.gaitpost.2017.12.026] [Medline: 29306811]
14. Girard J, Cohn J, Mahoor M, Mavadati S, Rosenwald D. Social Risk and Depression: Evidence from Manual and Automatic Facial Expression Analysis. Proc Int Conf Autom Face Gesture Recognit 2013:1-8 [FREE Full text] [doi: 10.1109/FG.2013.6553748] [Medline: 24598859]
15. Garcia-Toro M, Talavera JA, Saiz-Ruiz J, Gonzalez A. Prosody impairment in depression measured through acoustic analysis. J Nerv Ment Dis 2000 Dec;188(12):824-829. [doi: 10.1097/00005053-200012000-00006] [Medline: 11191583]
16. El Mansari M, Guiard BP, Chernoloz O, Ghanbari R, Katz N, Blier P. Relevance of norepinephrine-dopamine interactions in the treatment of major depressive disorder. CNS Neurosci Ther 2010 Jul;16(3):e1-17 [FREE Full text] [doi: 10.1111/j.1755-5949.2010.00146.x] [Medline: 20406250]
17. Herrera-Guzmán I, Gudayol-Ferré E, Herrera-Guzmán D, Guàrdia-Olmos J, Hinojosa-Calvo E, Herrera-Abarca JE. Effects of selective serotonin reuptake and dual serotonergic-noradrenergic reuptake treatments on memory and mental processing

speed in patients with major depressive disorder. J Psychiatr Res 2009 Jul;43(9):855-863. [doi: 10.1016/j.jpsychires.2008.10.015] [Medline: 19128810]

18. Morrissette DA, Stahl SM. Modulating the serotonin system in the treatment of major depressive disorder. CNS Spectr 2014 Dec;19 Suppl 1:57-67; quiz 54. [doi: 10.1017/S1092852914000613] [Medline: 25544378]
19. Stockmeier CA. Involvement of serotonin in depression: evidence from postmortem and imaging studies of serotonin receptors and the serotonin transporter. Journal of Psychiatric Research 2003 Sep;37(5):357-373. [doi: 10.1016/s0022-3956(03)00050-5] [Medline: 12849929]
20. Stockmeier CA, Shapiro LA, Dilley GE, Kolli TN, Friedman L, Rajkowska G. Increase in serotonin-1A autoreceptors in the midbrain of suicide victims with major depression-postmortem evidence for decreased serotonin activity. J Neurosci 1998 Oct 15;18(18):7394-7401 [FREE Full text] [Medline: 9736659]
21. The Lancet. SSRIs: suicide risk and withdrawal. The Lancet 2003 Jun 14;361(9374):1999. [doi: 10.1016/S0140-6736(03)13671-9] [Medline: 12814704]
22. Wessely S, Kerwin R. Suicide risk and the SSRIs. JAMA 2004 Jul 21;292(3):379-381. [doi: 10.1001/jama.292.3.379] [Medline: 15265855]
23. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. npj Digit. Med 2019 Mar 11;2(1):40 [FREE Full text] [doi: 10.1038/s41746-019-0090-4] [Medline: 31304386]
24. Torous J, Onnela J, Keshavan M. New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Transl Psychiatry 2017 Mar 07;7(3):e1053 [FREE Full text] [doi: 10.1038/tp.2017.25] [Medline: 28267146]
25. Lenze EJ, Rodebaugh TL, Nicol GE. A Framework for Advancing Precision Medicine in Clinical Trials for Mental Disorders. JAMA Psychiatry 2020 Jul 01;77(7):663-664. [doi: 10.1001/jamapsychiatry.2020.0114] [Medline: 32211837]
26. Kliem S, Lohmann A, Mößle T, Brähler E. German Beck Scale for Suicide Ideation (BSS): psychometric properties from a representative population survey. BMC Psychiatry 2017 Dec 04;17(1):389 [FREE Full text] [doi: 10.1186/s12888-017-1559-9] [Medline: 29202737]
27. Nock MK, Holmberg EB, Photos VI, Michel BD. Self-Injurious Thoughts and Behaviors Interview: development, reliability, and validity in an adolescent sample. Psychol Assess 2007 Oct;19(3):309-317. [doi: 10.1037/1040-3590.19.3.309] [Medline: 17845122]
28. Chu C, Hom MA, Stanley IH, Gai AR, Nock MK, Gutierrez PM, et al. Non-suicidal self-injury and suicidal thoughts and behaviors: A study of the explanatory roles of the interpersonal theory variables among military service members and veterans. J Consult Clin Psychol 2018 Jan;86(1):56-68 [FREE Full text] [doi: 10.1037/ccp0000262] [Medline: 29172592]
29. Brahmbhatt S. Introduction to Computer Vision and OpenCV. In: Practical OpenCV. Berkeley, CA: Apress; 2013:3-5.
30. Baltrusaitis T, Robinson P, Morency L. OpenFace: An open source facial behavior analysis toolkit. 2016 Presented at: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV); March 7-10, 2016; Lake Placid, NY. [doi: 10.1109/wacv.2016.7477553]
31. Ekman P, Rosenberg E. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford, United Kingdom: Oxford University Press; 2005.
32. Jadoul Y, Thompson B, de Boer B. Introducing Parselmouth: A Python interface to Praat. Journal of Phonetics 2018 Nov;71:1-15. [doi: 10.1016/j.wocn.2018.07.001]
33. Cohen ML. An Adaptive R-Estimate. Fort Belvoir, VA: Defense Technical Information Center; 1980.
34. Insel TR. Digital phenotyping: a global tool for psychiatry. World Psychiatry 2018 Oct;17(3):276-277 [FREE Full text] [doi: 10.1002/wps.20550] [Medline: 30192103]
35. Abbas A. anzarabbas/ms_dbms_mdd_suicidality. GitHub, Inc. URL: https://github.com/anzarabbas/ms_dbms_mdd_suicidality [accessed 2021-01-15]
36. Galatzer-Levy I, Abbas A, Yadav V, Koesmahargyo V, Aghjayan A, Marecki S, et al. Remote digital measurement of visual and auditory markers of Major Depressive Disorder severity and treatment response. medRxiv Preprint posted online August 26, 2020. [doi: 10.1101/2020.08.24.20178004]
37. Galatzer-Levy IR, Abbas A, Koesmahargyo V, Yadav V, Perez-Rodriguez MM, Rosenfield P, et al. Facial and vocal markers of schizophrenia measured using remote smartphone assessments. medRxiv Preprint posted online December 4, 2020. [doi: 10.1101/2020.12.02.20219741]
38. Abbas A, Yadav V, Smith E, Ramjas E, Rutter SB, Benavides C, et al. Computer vision-based assessment of motor functioning in schizophrenia: Use of smartphones for remote measurement of schizophrenia symptomatology. medRxiv Preprint posted online July 25, 2020. [doi: 10.1101/2020.07.20.20158287]
39. Abbas A, Schultebraucks K, Galatzer-Levy I. Digital Measurement of Mental Health: Challenges, Promises, and Future Directions. Psychiatric Annals 2021 Jan 01;51(1):14-20. [doi: 10.3928/00485713-20201207-01]
40. Insel TR. Digital phenotyping: a global tool for psychiatry. World Psychiatry 2018 Oct;17(3):276-277 [FREE Full text] [doi: 10.1002/wps.20550] [Medline: 30192103]

Abbreviations

BSS: Beck Scale for Suicide Ideation
MDD: major depressive disorder
SSRI: selective serotonin reuptake inhibitor

Edited by R Kukafka; submitted 26.10.20; peer-reviewed by A Benis, P Gazerani; comments to author 24.11.20; revised version received 15.12.20; accepted 16.03.21; published 03.06.21.

Please cite as:
Galatzer-Levy I, Abbas A, Ries A, Homan S, Sels L, Koesmahargyo V, Yadav V, Colla M, Scheerer H, Vetter S, Seifritz E, Scholz U, Kleim B
Validation of Visual and Auditory Digital Markers of Suicidality in Acutely Suicidal Psychiatric Inpatients: Proof-of-Concept Study
J Med Internet Res 2021;23(6):e25199
URL: https://www.jmir.org/2021/6/e25199
doi:10.2196/25199
PMID:34081022

Original Paper

Proof-of-Concept Support for the Development and Implementation of a Digital Assessment for Perinatal Mental Health: Mixed Methods Study

Nayra Anna Martin-Key[1*], PhD; Benedetta Spadaro[1*], MPhil; Thea Sofie Schei[2], PhD; Sabine Bahn[1], MD, PhD, MRCPsych, FRSB

[1]Cambridge Centre for Neuropsychiatric Research, Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, United Kingdom

[2]Psyomics Ltd, Cambridge, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Sabine Bahn, MD, PhD, MRCPsych, FRSB
Cambridge Centre for Neuropsychiatric Research
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom
Phone: 44 1223 334151
Email: sb209@cam.ac.uk

Abstract

Background: Perinatal mental health symptoms commonly remain underdiagnosed and undertreated in maternity care settings in the United Kingdom, with outbreaks of disease, like the COVID-19 pandemic, further disrupting access to adequate mental health support. Digital technologies may offer an innovative way to support the mental health needs of women and their families throughout the perinatal period, as well as assist midwives in the recognition of perinatal mental health concerns. However, little is known about the acceptability and perceived benefits and barriers to using such technologies.

Objective: The aim of this study was to conduct a mixed methods evaluation of the current state of perinatal mental health care provision in the United Kingdom, as well as users' (women and partners) and midwives' interest in using a digital mental health assessment throughout the perinatal period.

Methods: Women, partners, and midwives were recruited to participate in the study, which entailed completing an online survey. Quantitative data were explored using descriptive statistics. Open-ended response data were first investigated using thematic analysis. Resultant themes were then mapped onto the components of the Capability, Opportunity, and Motivation Behavior model and summarized using descriptive statistics.

Results: A total of 829 women, 103 partners, and 90 midwives participated in the study. The provision of adequate perinatal mental health care support was limited, with experiences varying significantly across respondents. There was a strong interest in using a digital mental health assessment to screen, diagnose, and triage perinatal mental health concerns, particularly among women and midwives. The majority of respondents (n=781, 76.42%) expressed that they would feel comfortable or very comfortable using or recommending a digital mental health assessment. The majority of women and partners showed a preference for in-person consultations (n=417, 44.74%), followed by a blended care approach (ie, both in-person and online consultations) (n=362, 38.84%), with fewer participants preferring online-only consultations (n=120, 12.88%). Identified benefits and barriers mainly related to physical opportunity (eg, accessibility), psychological capability (eg, cognitive skills), and automatic motivation (eg, emotions).

Conclusions: This study provides proof-of-concept support for the development and implementation of a digital mental health assessment to inform clinical decision making in the assessment of perinatal mental health concerns in the United Kingdom.

(J Med Internet Res 2021;23(6):e27132) doi:10.2196/27132

KEYWORDS

COM-B; COVID-19; digital mental health; maternal mental health; paternal mental health; perinatal mental health; mental health; support; development; implementation; assessment; mother; women

Introduction

The perinatal period, comprising pregnancy along with one year after giving birth, represents one of the most complex and challenging times in a woman's life, with perinatal mental health disorders affecting around 15%-20% of women in the general population [1] and up to 40% of women in intensive perinatal care units [2]. Common psychiatric complications during the perinatal period include depression and anxiety [3]. Other less common conditions include new onset or recurring bipolar disorder, schizophrenia, and other psychotic illnesses, with most psychotic episodes occurring within the first 2 postnatal weeks [4]. Mental health concerns during the perinatal period do not only affect women but can also affect their partners, with approximately 5%-10% of fathers experiencing depression [5] and 5%-15% experiencing anxiety [6].

Despite these figures, perinatal mental health concerns commonly remain underdiagnosed and undertreated in maternity care settings in the United Kingdom [7]. This is likely due to a variety of reasons, including stigma and discomfort about discussing one's own mental health, as well as a fear of consequences of disclosure [8]. Other challenges relate to a lack of mental health training and formal education, particularly among midwives, time constraints, a pressure to prioritize physical over mental health [9], and a lack of knowledge regarding available referral pathways [10].

The current long-term costs of unidentified perinatal mental health concerns to society are estimated at around £8.1 billion (US $10.1 billion) for each 1-year cohort of births in the United Kingdom, of which 72% relate to adverse impacts on the child [7]. Critically, it is estimated that this figure will be astronomically higher due to the ongoing COVID-19 pandemic. Indeed, preliminary research on COVID-19 has indicated a significant increase in psychological distress for expectant mothers [11], with strict lockdown measures disrupting routine clinical appointments and access to mental health services, leaving many at an increased risk of poor mental health [12]. These measures have also led to increased financial difficulties, higher rates of domestic violence, as well as impaired support from family and friends, all of which have been identified as risk factors for perinatal mental health difficulties [13].

In light of the recently published recommendations by the Royal College of Obstetricians and Gynaecologists [13] on the use of remote means to provide support to women throughout the perinatal period, digital technologies may offer an innovative way to support the mental health needs of women and their families throughout and beyond these trying times. Notably, digital technologies have the potential to support midwives in the recognition of perinatal mental health concerns and patients' treatment needs. Further, research has demonstrated that individuals are more likely to report severe symptoms on technology platforms than to a health care professional (HCP) [14], and patients value the independence and empowerment that can be obtained via the use of a digital platform [15]. However, little is known about the acceptability and perceived benefits and barriers to using digital technologies in perinatal mental health care, particularly among women and partners at different stages of parenthood, as well as midwives.

The present mixed methods study aimed to examine the current state of perinatal mental health care provision in the United Kingdom, as well as in the context of the evolving pandemic, relative to the recommendations put forward by the National Institute for Health and Care Excellence (NICE) [1]. Further, we explored users' (women and partners) and midwives' attitudes toward using a digital mental health assessment to screen, diagnose, and triage perinatal mental health concerns. To this end, we analyzed participants' attitudes in light of the Capability, Opportunity, and Motivation Behavior (COM-B) model, a theory of behavior that poses behavior as a result of the interaction between distinct behavioral components, namely capability, opportunity, and motivation [16]. The COM-B model has been commonly employed to design health care interventions and understand complex behaviors that result from the interplay between different stakeholders (eg, patients, carers, and health care providers) [17-19]. In our study, the model acts as a theoretical framework, informing future strategies that can address the barriers and facilitators to the uptake of a digital mental health assessment.

The ultimate goal of this research was to provide proof-of-concept support for the development and implementation of a digital mental health assessment tool to inform clinical decision making regarding the diagnosis and treatment of perinatal mental health concerns in the United Kingdom.

Methods

Participants

Participants were recruited between April and August 2020 via email, posts on the Facebook and Twitter pages of the Cambridge Centre for Neuropsychiatric Research, and paid Facebook advertisements. All participants provided informed consent electronically to participate in the study, which was approved by the University of Cambridge Human Biology Research Ethics Committee (approval number PRE.2020.041). Inclusion criteria for the study were: (1) ≥18 years, (2) UK residence, and (3) fluency in English. Midwives were also required to be currently practicing in the United Kingdom, while women and partners were required to fall into one of the following subgroups: (1) currently planning or trying to conceive, (2) currently pregnant or partner of someone who is currently pregnant, or (3) given birth within the last 2 years or partner of someone who has given birth within the last 2 years. There were no other inclusion criteria.

The women and partners were invited to enter their email for the chance to win a £50 (US $67) Highstreet voucher, while midwives were provided with a £15 (US $20) Highstreet voucher for their time. Of the partners who participated in the study, 65.05% (n=67) were recruited via participating women, while midwives were recruited separately, meaning that they were not necessarily involved in the care of the participating women.

Materials and Procedure

Three anonymous online surveys were created using Qualtrics in order to explore the current state of perinatal mental health provision in the United Kingdom, as well as attitudes toward using a digital mental health assessment to screen, diagnose, and triage perinatal mental health concerns. All questions were developed in consultation with the senior author (SB), a practicing psychiatrist. The surveys were adaptive in nature, such that only relevant questions were asked based on previous responses.

The women and partner surveys comprised two separate surveys that could be completed in 10-15 minutes and included five sections: (1) sociodemographic information, (2) perinatal health information, (3) mental health care provision, (4) mental health symptoms, (5) interest in a digital mental health assessment, and (6) perceived benefits and barriers to using a digital mental health assessment. The women's survey also included an additional section on COVID-19 and mental health (only data from women who reported being pregnant when completing the survey or who had given birth within the last 3 months and had been in contact with a midwife since the United Kingdom entered its first lockdown [end of March to beginning of July 2020] were collected). The questions were assessed for relevance, appropriateness, and length by 2 women and 2 partners. Table S1 in Multimedia Appendix 1 provides the questions included in the surveys.

The midwives' survey could be completed in 20-30 minutes and comprised five sections: (1) sociodemographic information, (2) mental health provision, (3) partners' mental health care provision, (4) COVID-19 and perinatal mental health care provision, (5) interest in a digital mental health assessment, and (6) perceived benefits and barriers to using a digital mental health assessment. The questions were assessed for relevance, appropriateness, and length by a practicing midwife and a practicing specialist perinatal psychiatrist. Table S1 in Multimedia Appendix 2 provides the questions included in the survey.

Data Analytic Strategy

The processing and analysis of quantitative data (ie, frequencies and percentages) was conducted in R, version 4.0.2 (R Foundation for Statistical Computing) [20]. Figures were created using the R packages *ggplot2* and *likert*, versions 3.3.2 and 1.3.5, respectively, as well as Excel, version 16.43 (Microsoft Corp). These data included sociodemographic information, perinatal health characteristics, mental health provision, mental health symptoms, mental health and COVID-19, as well as attitudes toward the use of a digital mental health assessment throughout the perinatal period.

Qualitative data were analyzed following the recommended six stages of thematic analysis [21]. These data comprised open-ended responses regarding the perceived benefits and barriers to using a digital mental health assessment throughout the perinatal period. The first author (NAM-K) became familiar with the data by reading and rereading the open-ended responses and noting down ideas. Initial codes were then generated, allocated, and incorporated into a coding scheme, encompassing brief descriptions for each code. The second author (BS) performed a blinded allocation of codes to each of the open-ended responses using the coding scheme. Following this, any inconsistencies were discussed until a consensus was reached. The codes were then condensed into broad themes, with the first and second authors generating themes separately. These were then discussed until a consensus was reached.

The COM-B [16] was then used to condense themes further into the model's framework and components, namely: Capacity (components: physical and psychological capacity), Opportunity (components: physical and social opportunity), and Motivation (components: automatic and reflective motivation). The COM-B model proposes that individuals need capability (C), opportunity (O) and motivation (M) to engage in a particular behavior (B). In the model, capability is defined as an individual's ability to physically and psychologically perform a behavior of interest, opportunity indicates the external factors that can affect engagement in said behavior, and motivation refers to a construct comprising automatic and reflective drives. This model has been commonly used to design interventions and understand behaviors in clinical and public health contexts (eg, [22-26]). To map themes onto the COM-B model, both the first and second authors allocated a COM-B component to each of the established themes under blinded conditions. Once again, any inconsistencies were discussed until a consensus was reached. Frequencies were calculated per group status (ie, women, partners, midwives), with between-group comparisons assessed using the Fisher exact test. Post hoc pairwise comparisons (ie, women vs partners, women vs midwives, and partners vs midwives) were conducted using the Fisher exact test where appropriate.

Results

Sociodemographic Characteristics

A total of 829 women (planning or trying to conceive: n=76; currently pregnant: n=259; had given birth within the last 2 years: n=494), 103 partners (planning or trying to conceive: n=11; partner of someone who is currently pregnant: n=38; partner of someone who has given birth within the last 2 years: n=54), and 90 midwives participated in the study.

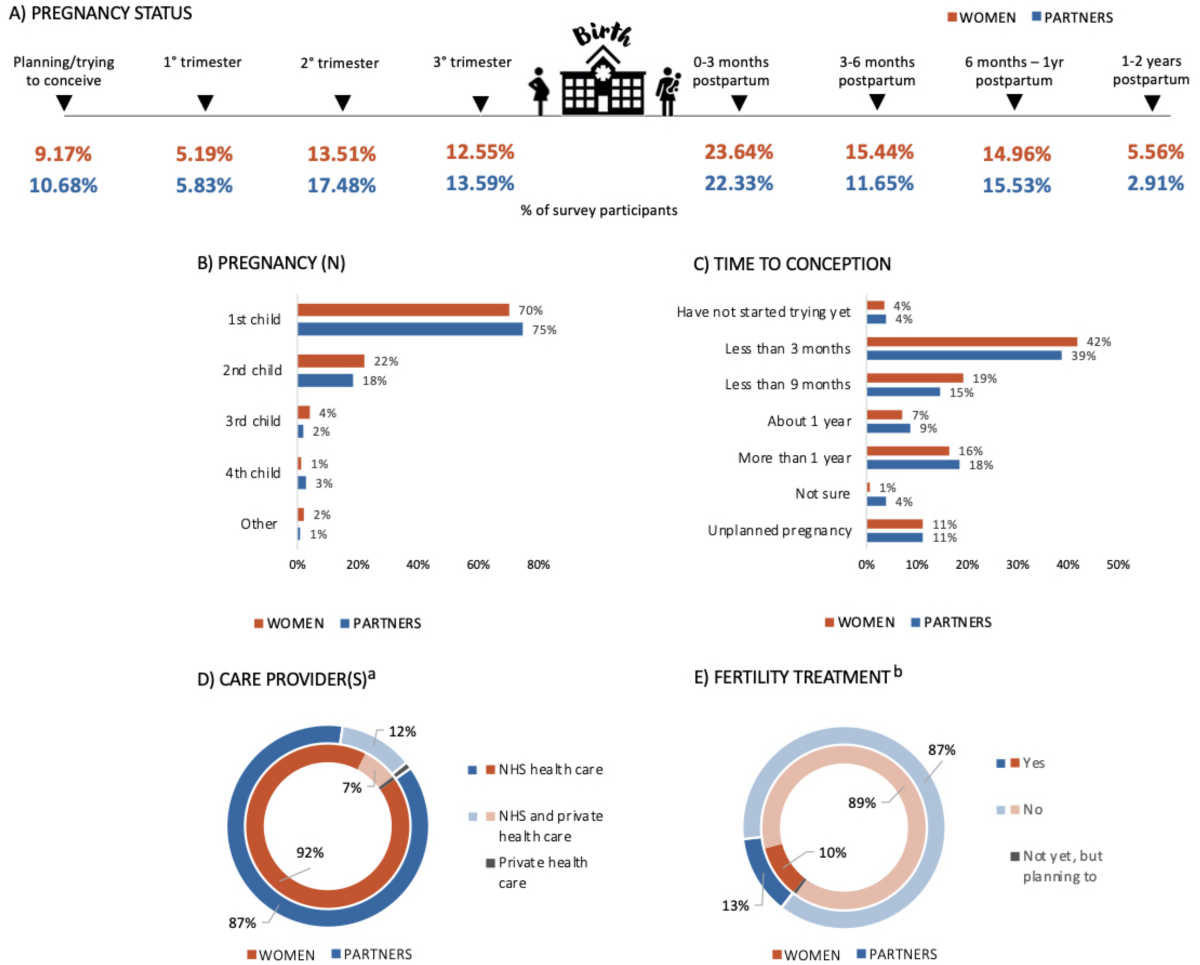
Women's and partners' sociodemographic information can be found in Table S1A (Multimedia Appendix 1). The average ages of the women and partners were 31.78 (SD 4.63) years and 34.87 (SD 6.38) years, respectively, with the majority of respondents being White. Over 80% (n=687) of women and approximately 75% (n=77) of partners had at least an undergraduate degree, and the majority of respondents were married or in a civil partnership. Across the women and partner groups, around 71% (n=659) owned their home and approximately 78% (n=724) had a household income of at least £35,001 (US $46,950) before tax.

Midwives' sociodemographic information can be found in Table S1A (Multimedia Appendix 2). The average age of the midwives was 39.90 (SD 9.67) years, with the majority being female and having an undergraduate degree. The average number of years of practice was 10.36 (SD 7.19) years.

Perinatal Health Characteristics

Perinatal health characteristics for women and partners can be found in Figure 1 and Table S1B (Multimedia Appendix 1). Over half of the respondents had given birth or had a partner who had given birth within the last 2 years (Figure 1A). For the majority of respondents, this was or would be their first child (Figure 1B) and conception had occurred within 1 year (Figure 1C). The vast majority of respondents' pregnancy or partner's pregnancy was solely monitored by the National Health Service (NHS) (Figure 1D), and the majority of respondents had not received any fertility treatment (Figure 1E). A small percentage of women (n=49, 6.93%) and partners (n=14, 16.09%) expressed having had to terminate the pregnancy (eg, abortion, ectopic pregnancy, other medical intervention). Just under a quarter of respondents had a miscarriage (n=157, 22.21%) or had a partner who had a miscarriage (n=16, 18.39%), while around half of the respondents had experienced a difficult birth (n=262, 53.04%) or had a partner who had experienced a difficult birth (n=26, 48.15%).

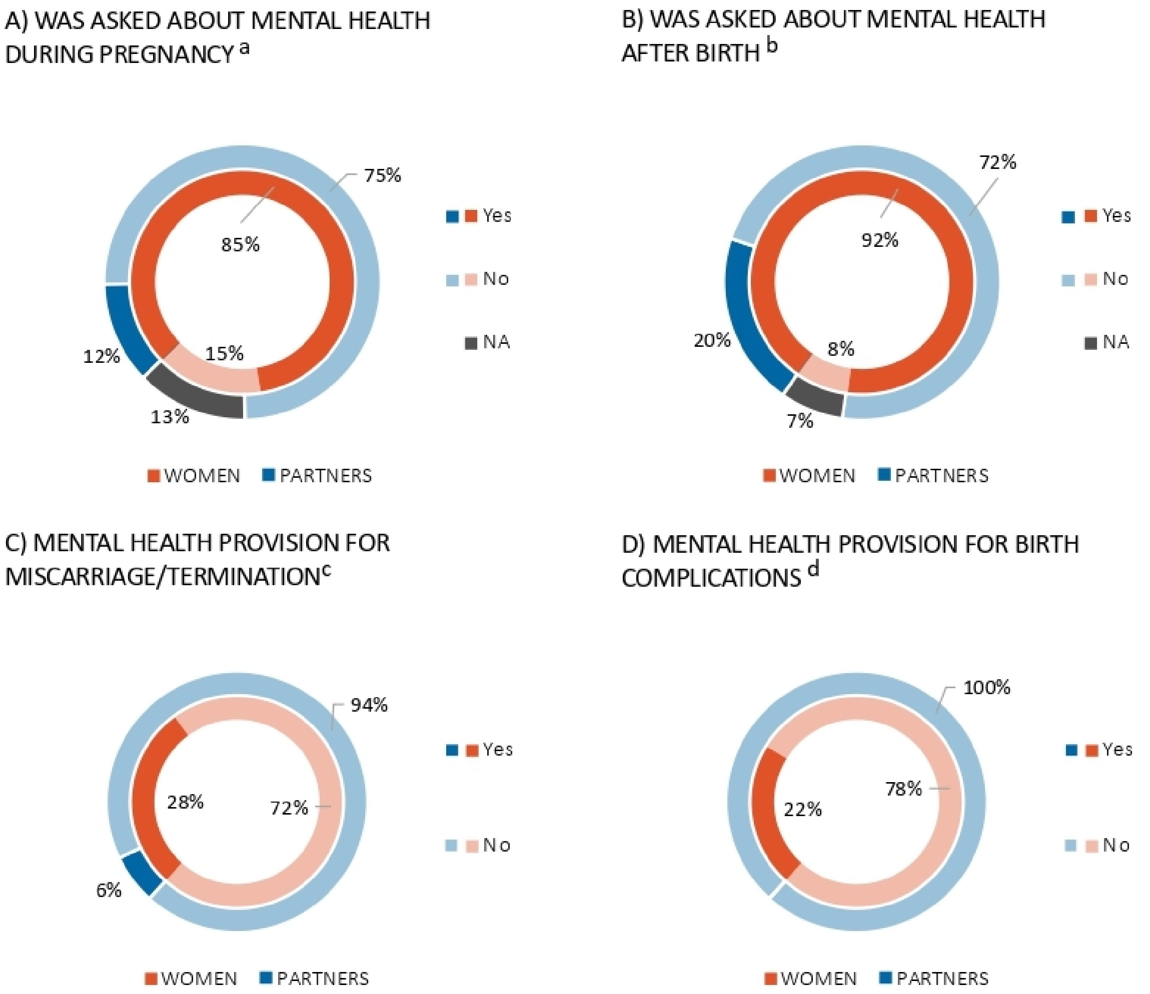
Figure 1. Perinatal health characteristics of women and partners, including pregnancy status (A), pregnancy number (B), time to conception (C), care provider(s) (D), and fertility treatment (E). a: Includes those in contact with the health care system (women: n=785; partners: n=93); b: Includes those who have started trying to conceive and planned pregnancies (women: n=707; partners: n=87). NHS: National Health Service.



Mental Health Care Provision

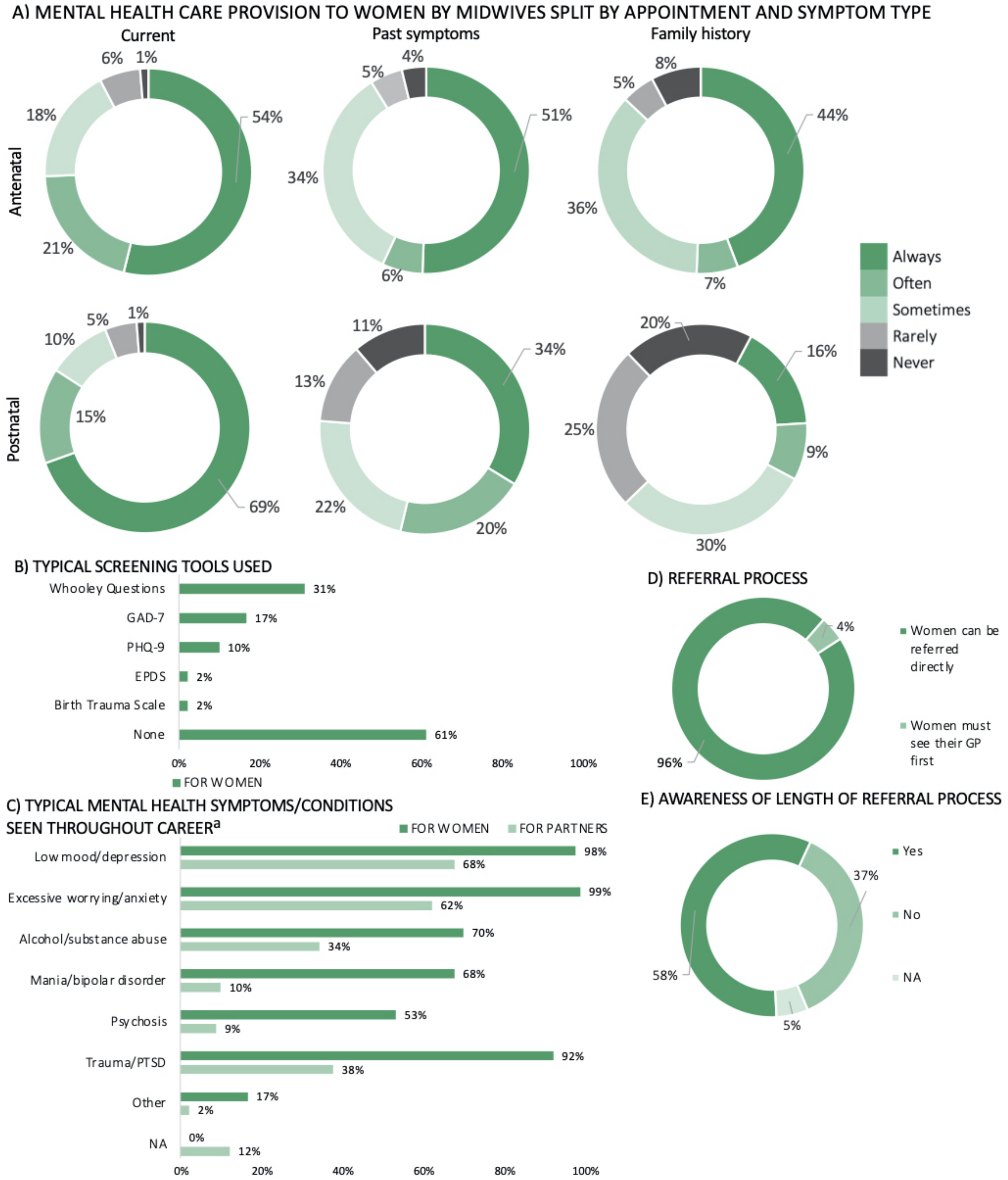
Women's and partners' reported experiences with mental health care provision throughout the perinatal period can be found in Figure 2 and Table S1C (Multimedia Appendix 1). Over two-thirds of women (n=576, 73.38%) and 16.13% (n=15) of partners had received information on mental health during antenatal or postnatal appointments, with information typically provided in the form of face-to-face discussions and leaflets. The majority of women were asked about their mood or mental health at least once throughout their pregnancy or after giving birth, while the opposite was the case for partners (Figure 2A and B). Just over a quarter of women were offered mental health support or advice following a miscarriage and/or termination (Figure 2C), while 22.14% (n=58) received mental health support or advice after a difficult birth (Figure 2D). Only 1 partner (6.25%) was offered mental health support or advice following their partner's miscarriage or termination, and none were offered any mental health support or advice after their partner's difficult birth (Figure 2C and D).

Figure 2. Perinatal mental health provision for women and partners, including being asked about their mental health during pregnancy (A) and after giving birth (B), as well as being offered mental health support following a miscarriage or termination (C) and after a difficult birth (D). a: Includes those who are pregnant or have given birth (women: n=753; partners: n=92); b: Includes those who have given birth (women: n=494; partners: n=54); c: Includes those who answered "yes" to the termination or miscarriage question (women: n=159; partners: n=16); d: Includes those who answered "yes" to birth complications question (women: n=262; partners: n=26). NA: not applicable.



A summary of midwives' experiences providing perinatal mental health care can be found in Figure 3 and Table S1B and S1C (Multimedia Appendix 2). One-third of midwives (n=30, 33.33%) reported always providing women with information on mental health during antenatal or postnatal appointments, with information typically provided during face-to-face appointments and via leaflets (Table S1B, Multimedia Appendix 2). Approximately 50% (n=42) of midwives reported always asking patients about their current mental health symptoms throughout the antenatal period, while almost two-thirds (n=57) expressed always asking patients about their current mental health symptoms throughout the postnatal period (Figure 3A). These figures were lower when inquiring about past symptoms and a family history of mental health, throughout both the antenatal and postnatal periods (Figure 3A). The vast majority of midwives (n=55, 61.11%) used nonstandardized questions to assess patients' mental health symptoms, while just under one-third (n=28, 31.11%) used the Whooley Questions, and fewer used tools such as the Patient Health Questionnaire-9 (PHQ-9) [27] (n=9, 10.00%) and the 7-item Generalized Anxiety Disorder questionnaire (GAD-7) [28] (n=15, 16.67%) (Figure 3B). The most common mental health conditions encountered were depression, anxiety, and trauma or posttraumatic stress disorder (PTSD) (Figure 3C). The majority of midwives (n=86, 95.56%) reported being able to directly refer a patient to a mental health specialist, with just over a half (n=52, 57.78%) being aware of the length of the referral process (Figure 3D and E).

Figure 3. Midwives' provision of perinatal mental health care, including mental health care provision provided to women split by appointment type (ie, antenatal and postnatal) and symptom type (ie, current, past, and family history) (A), typical mental health screening tools used (B), typical mental health symptoms or conditions encountered throughout their career (C), referral process (ie, whether patients can be directly referred to a mental health specialist or whether they have to see their general practitioner first (GP) (D), and awareness of the referral process (E). a: Percentages add to more than 100% as midwives could select multiple options. GAD-7: Generalized Anxiety Disorder-7, PHQ: Patient Health Questionnaire, EPDS: Edinburgh Postnatal Depression Scale, PTSD: posttraumatic stress disorder, NA: not applicable.

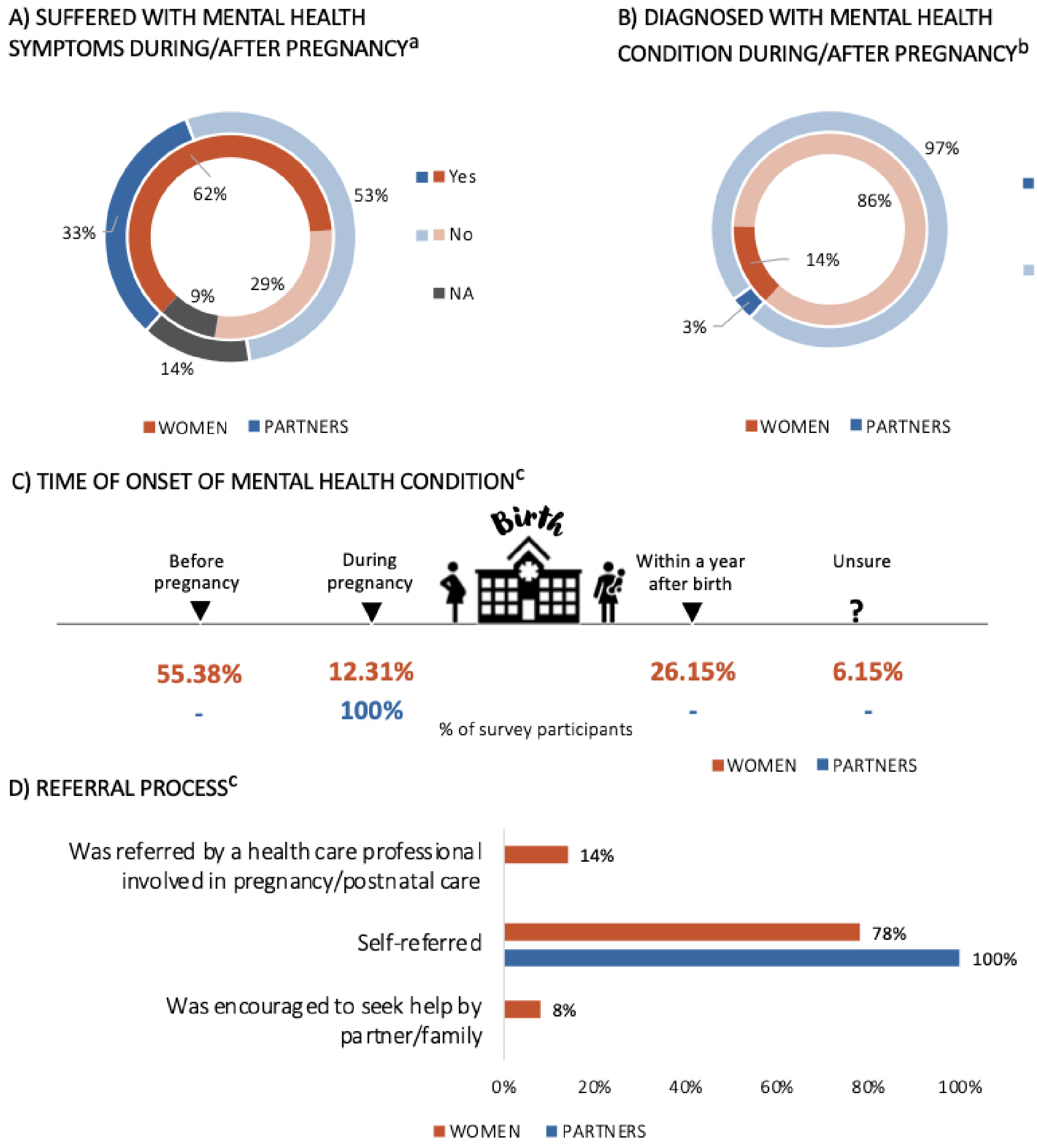


Approximately 7% (n=6) of midwives reported always providing partners with information on mental health (Table S1C, Multimedia Appendix 2). When provided, mental health information was typically made available during face-to-face appointments (n=56, 87.50%) (Table S1C, Multimedia Appendix 2). Similar to the women, the most common mental health conditions seen in partners were depression, anxiety, and trauma or PTSD (Figure 3C).

Mental Health Symptoms and Diagnosis

A summary of the women's and partners' mental health symptoms and diagnoses throughout the perinatal period can be found in Figure 4 and Table S1D (Multimedia Appendix 1). Almost two-thirds of women (n=469, 62.28%) and one-third of partners (n=30, 32.61%) reported having experienced mental health symptoms during the pregnancy and/or after delivery (Figure 4A), with approximately 13.86% (n=65) of women and around 3.33% (n=1) of partners being diagnosed with a mental health condition by a health care professional (eg, general practitioner or psychiatrist) during the perinatal period (Figure 4B). The most common diagnoses were depression and anxiety (Table S1D, Multimedia Appendix 1). Over one half of women (n=36, 55.38%) reported having had the condition before they became pregnant, while 26.15% (n=17) developed the condition within 1 year after giving birth (Figure 4C). The vast majority of respondents had referred themselves to a mental health specialist, while approximately 14% of women were referred to a mental health professional by a midwife or HCP involved in their antenatal or postnatal care (Figure 4D).

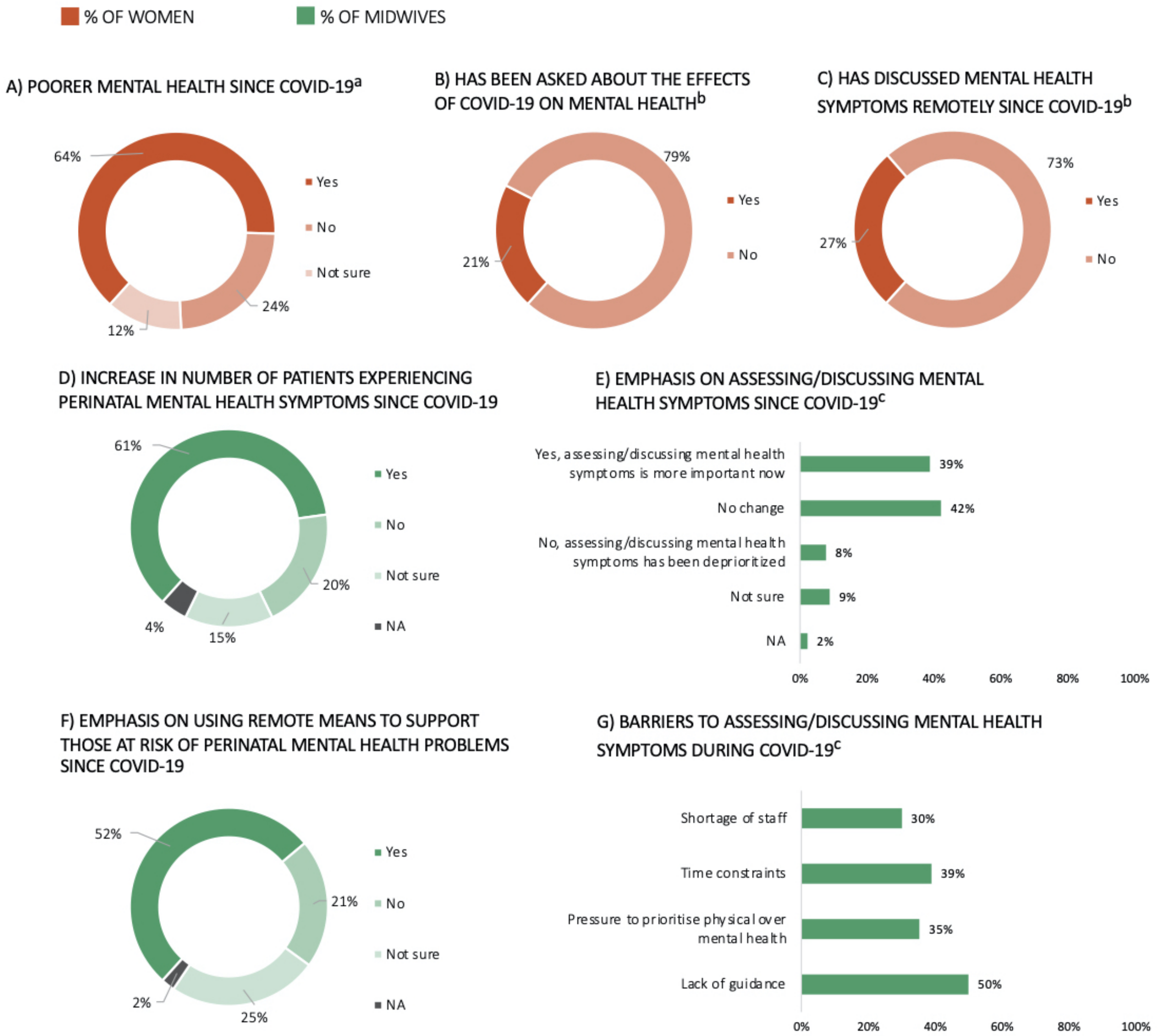
Figure 4. Mental health symptoms and diagnoses for women and partners, including whether they had experienced mental health symptoms during or after the pregnancy (A), whether they had been diagnosed with a mental health condition during or after the pregnancy (B), the time of onset of the mental health condition (C), and information regarding the referral process (D). a: Includes those who are pregnant or have given birth (women: n=753; partners: n=92); b: Includes those who answered "yes" to experiencing mental health symptoms during or after pregnancy (women: n=469; partners: n=30); c: Includes those who were diagnosed with a mental health condition (women: n=65; partners: n=1). NA: not applicable.



COVID-19 and Mental Health

A summary of the effects of COVID-19 on women's mental health can be found in Figure 5 and Table S1E (Multimedia Appendix 1). Almost two-thirds of women (n=293, 64.40%) reported poorer mental health symptoms following the COVID-19 outbreak (Figure 5A). Approximately 21% of women (n=93) were specifically asked about the effects of the evolving pandemic on their mental health (Figure 5B), while nearly 27% (n=120) discussed their mental health remotely (ie, via telephone or video consultation) with a midwife or HCP involved in their antenatal or postnatal care (Figure 5C).

Figure 5. The effects of COVID-19 on women’s mental health and midwives’ mental health care provision, including the extent to which women had experienced poorer mental health since the outbreak (A), whether they had been asked about the effects of the pandemic on their mental health (B), and whether remote means (ie, telephone or video consultations) had been used to discuss their mental health throughout the pandemic (C). Regarding midwives’ mental health care provision, this included the extent to which midwives had seen an increase in the number of patients experiencing mental health difficulties since COVID-19 (D), whether there had been an emphasis on assessing or discussing mental health symptoms throughout the pandemic (E), whether there had been an emphasis on using remote means (ie, telephone or video consultations) to support those at risk of perinatal mental health difficulties since COVID-19 (F), and the barriers to assessing or discussing mental health difficulties throughout the pandemic (G). a: Includes women who are pregnant or have given birth within the last 3 months (n=455); b: Includes women who are pregnant or have given birth within the last 3 months and have been in contact with a midwife since the lockdown (n=446); c: Percentages add to more than 100% as midwives could select multiple options. NA: not applicable.

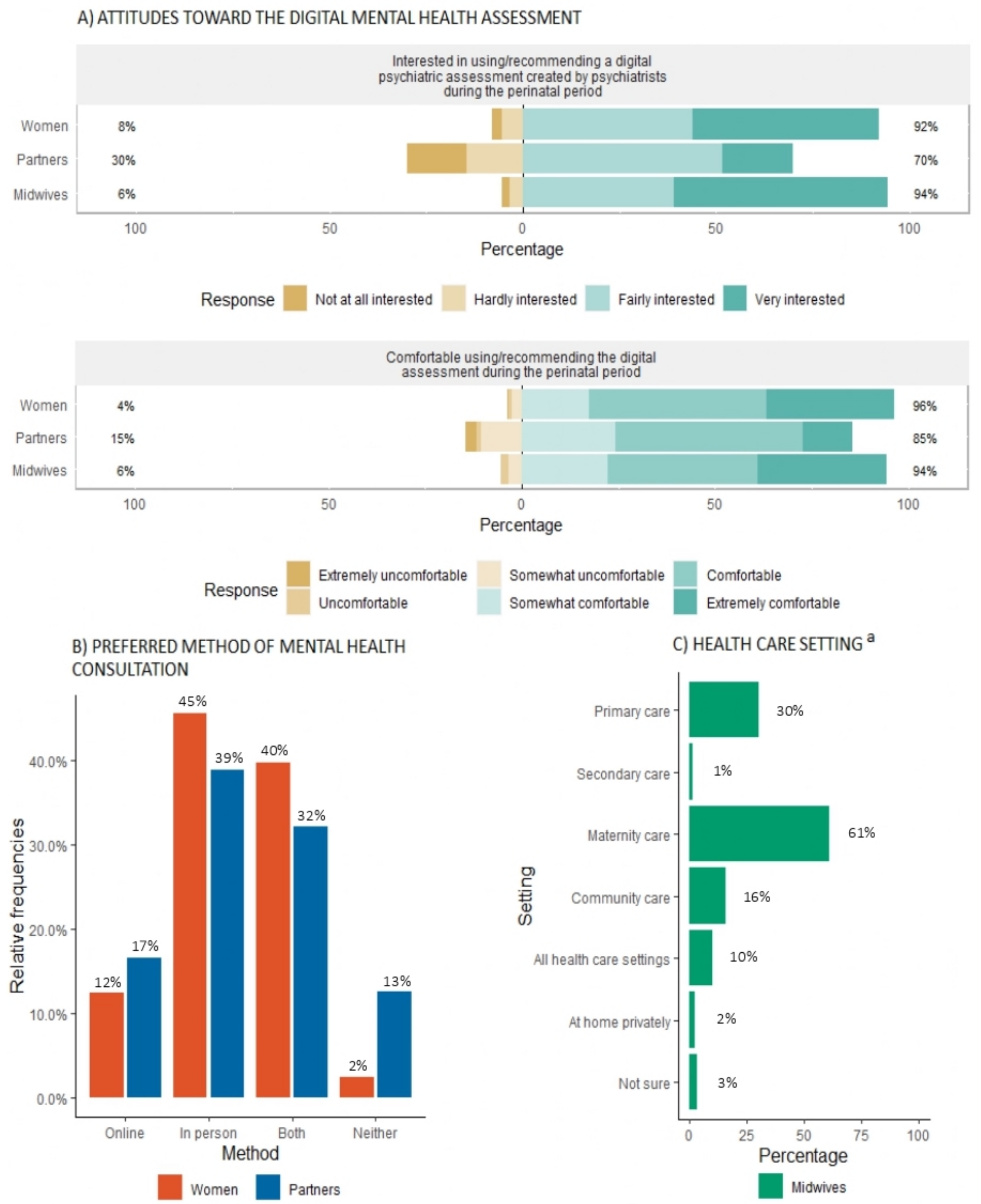


The effects of COVID-19 on midwives’ experiences providing mental health care are summarized in Figure 5 and Table S1D in Multimedia Appendix 2. A total of 55 (61.11%) midwives reported seeing an increase in perinatal mental health symptoms since the COVID-19 outbreak (Figure 5D). The majority of midwives reported there being no change to the standard of mental health care, although this was closely followed by an emphasis on prioritizing asking about mental health (Figure 5E). Further, 52.22% (n=47) of respondents reported using remote means (ie, telephone or video consultation) to discuss patients’ mental health symptoms (Figure 5F), and 50% (n=45) of midwives saw a lack of guidance on how to evaluate mental health symptoms throughout the pandemic as an important barrier to assessing perinatal mental health. Other important barriers were time constraints and a pressure to prioritize physical over mental health (Figure 5G).

Interest in a Digital Mental Health Assessment to Screen, Diagnose, and Triage Perinatal Mental Health Symptoms

Overall, there was a strong interest in using or recommending a digital mental health assessment among women and midwives, respectively, and the majority of respondents (n=781, 76.42%) expressed that they would feel comfortable or very comfortable using or recommending a digital mental health assessment (Figure 6A). The majority of women and partners showed a preference for in-person consultations (n=417, 44.74%), followed by a blended care approach (ie, both in-person and online consultations) (n=362, 38.84%), with fewer patients preferring online-only consultations (n=120, 12.88%) (Figure 6B). The vast majority of midwives (n=55, 61.11%) reported seeing the digital mental health assessment being best placed in maternity care settings (ie, antenatal and postnatal care) (Figure 6C).

Figure 6. Interest in a digital mental health assessment to screen, diagnose, and triage perinatal mental health symptoms, including respondents' interest in using or recommending a digital mental health assessment throughout the perinatal period (A), women's and partners' preferred method of mental health consultation following a suggestion by the digital mental health assessment that a follow-up assessment is recommended (B), and midwives' responses as to where they see the digital mental health assessment working best (C). a: Percentages add to more than 100% as some midwives suggested more than one health care setting.



Benefits and Barriers to Using a Digital Mental Health Assessment to Screen, Diagnose, and Triage Perinatal Mental Health Symptoms

The thematic analysis revealed 14 benefits and 17 barriers to using a digital mental health assessment to screen, diagnose, and triage perinatal mental health symptoms (see Table S1 in Multimedia Appendix 3 for initial codes, including how these were condensed into broad themes). Themes were then mapped onto the components of the COM-B model and are described below. Descriptive statistics and between-group comparisons per theme can be found in Table 1.

Table 1. Frequencies and between-group comparisons of the identified benefits and barriers to using a digital mental health assessment to screen, diagnose, and triage perinatal mental health symptoms.

COM-B[a] framework, components, and themes	Frequencies (n, %)[b]				*P* value[c]	Post hoc[d] analysis
	Overall	Women	Partners	Midwives		
Benefits						
Capability						
Physical capability						
Accuracy	42 (4.22)	34 (4.18)	5 (5.38)	3 (3.37)	.82	—[e]
Usability	102 (10.24)	78 (9.58)	12 (12.90)	12 (13.48)	.31	—
Psychological capability						
Cognitive skills	78 (7.83)	52 (6.39)	12 (12.90)	14 (15.73)	*.001*	P,M>W
Opportunity						
Physical opportunity						
Accessibility	303 (30.42)	261 (32.06)	20 (21.51)	22 (24.72)	.053	—
Support	193 (19.38)	139 (17.08)	18 (19.35)	36 (40.45)	*<.001*	M>W,P
Environment	483 (48.49)	417 (51.23)	39 (41.94)	27 (30.34)	*<.001*	W>M
Funding and implementation	23 (2.31)	19 (2.33)	3 (3.23)	1 (1.12)	.60	—
Social opportunity						
Peer support	10 (1.00)	6 (0.74)	2 (2.15)	2 (2.25)	.14	—
Normalization of mental health	93 (9.34)	84 (10.32)	3 (3.23)	6 (6.74)	*.05*	W>P
Motivation						
Automatic motivation						
Positive affect	204 (20.48)	176 (21.62)	14 (15.05)	14 (15.73)	.19	—
Reflective motivation						
Control	73 (7.33)	57 (7.00)	5 (5.38)	11 (12.36)	.15	—
Reassurance	45 (4.52)	36 (4.42)	7 (7.53)	2 (2.25)	.24	—
Honesty	55 (5.52)	43 (5.28)	3 (3.23)	9 (10.11)	.13	—
Beliefs	10 (1.00)	5 (0.61)	(0) 0.00	5 (5.62)	*.002*	M>W,P
None	2 (0.20)	1 (0.12)	0 (0.00)	1 (1.12)	.18	—
Barriers						
Capability						
Physical capability						
Accuracy	98 (10.25)	88 (11.41)	7 (7.29)	3 (3.37)	*.03*	W>M
Data protection	56 (5.86)	47 (6.10)	4 (4.17)	5 (5.62)	.84	—
Usability	25 (2.62)	20 (2.59)	3 (3.13)	2 (2.25)	.93	—
Psychological capability						
Knowledge	85 (8.89)	66 (8.56)	13 (13.54)	6 (6.74)	.21	—
Cognitive skills	168 (17.57)	123 (15.95)	17 (17.71)	28 (31.46)	*.002*	M>W,P
Technical skills	43 (4.50)	33 (4.28)	4 (4.17)	6 (6.74)	.51	—
Opportunity						
Physical opportunity						
Accessibility	180 (18.83)	138 (17.90)	10 (10.42)	32 (35.96)	*<.001*	M>W,P

COM-B[a] framework, components, and themes	Frequencies (n, %)[b]				*P* value[c]	Post hoc[d] analysis
	Overall	Women	Partners	Midwives		
Support	25 (2.62)	21 (2.72)	0 (0.00)	4 (4.49)	.10	—
Environment	198 (20.08)	154 (19.97)	20 (20.83)	18 (20.22)	.98	—
Funding and implementation	76 (7.95)	54 (7.00)	11 (11.46)	11 (12.36)	.08	—
Social opportunity						
Stigma	43 (4.50)	37 (4.80)	2 (2.08)	4 (4.49)	.56	—
Motivation						
Automatic motivation						
Negative affect	27 (2.82)	24 (3.11)	1 (1.04)	2 (2.25)	.62	—
Reflective motivation						
Control	66 (6.90)	55 (7.13)	8 (8.33)	3 (3.37)	.36	—
Beliefs	46 (4.81)	35 (4.54)	10 (10.42)	1 (1.12)	*.01*	P>W,M
Dishonesty	34 (3.56)	32 (4.15)	1 (1.04)	1 (1.12)	.17	—
Impersonal	126 (13.18)	106 (13.75)	11 (11.46)	9 (10.11)	.62	—
Reluctance	25 (2.62)	14 (1.82)	7 (7.29)	4 (4.49)	*.005*	P>W
None	60 (6.28)	47 (6.10)	8 (8.33)	5 (5.62)	.65	—

[a] COM-B: Capability, Opportunity, and Motivation Behavior model.

[b] Sample sizes varied for the benefits and barriers due to nonresponses (ie, no response was provided). Percentages were calculated based on the total number of responses, including no identified benefits or barriers (labeled as "none"). Benefits: overall (n=996), women (n=814), partners (n=93), midwives (n=89). Barriers: overall (n=956), women (n=771), partners (n=96), midwives (n=89).

[c] *P* values are based on the Fisher exact test. Italics indicates significant values.

[d] Pairwise comparisons; W: women, P: partners, M: midwives.

[e] Not applicable.

Physical Capability

Overall, just under 5% (n=42) of respondents described the digital mental health assessment as having the potential to provide more accurate diagnoses by facilitating clinical decision making and being more comprehensive and objective than the current standard of care. On the other hand, the complexity of mental health and the fact that a remote approach would impede users from delving deeper into particular symptoms or concerns were seen as issues that could affect the accuracy of the tool, with significantly more women (n=88, 11.41%) than midwives (n=3, 3.37%) identifying this as a barrier. Overall, usability (ie, how user-friendly the tool is to use) was also seen as a benefit (n=102, 10.24%) and a barrier (n=25, 2.59%), while just under 6% (n=56) of all respondents saw issues surrounding data protection as barriers.

Psychological Capability

Cognitive skills were regarded as both a benefit and a barrier. For instance, respondents saw the potential for increased awareness and understanding of mental health as a benefit to using a digital mental health assessment, with partners (n=12, 12.90%) and midwives (n=14, 15.73%) identifying this as a benefit to a significantly greater extent in comparison to women (n=52, 6.39%). On the other hand, one's capacity for self-reflection and awareness of difficulties, as well as general issues surrounding comprehension, were seen as barriers, with a significantly higher proportion of midwives (n=28, 31.46%) expressing these concerns relative to both the women (n=123, 15.95%) and partner (n=17, 17.71%) groups. Another barrier that was identified by approximately 9% (n=85) of all respondents regarded knowledge concerning the tool's existence, including a general lack of awareness regarding its usefulness. Furthermore, just under 5% (n=43) of all respondents regarded poor technical skills as a barrier to using a digital mental health assessment.

Social Opportunity

The normalization of mental health was regarded as a benefit to using the tool, with significantly more women (n=84, 10.32%) identifying this as a benefit relative to partners (n=3, 3.23%). Peer support, including the potential to connect with other users or to share one's experience with one's partner and/or family, was seen as a benefit seen by a small number of respondents (n=10, 1.00%). On the other hand, stigma surrounding the use of a digital mental health assessment was seen as a potential barrier by just under 5% (n=43) of all respondents.

Physical Opportunity

Accessibility was seen as both a benefit and a barrier. For instance, improved access to mental health care, including it being wide-reaching and inclusive, was regarded a benefit by

30.42% (n=303) of all respondents. On the contrary, living in rural or deprived areas, as well as not having access to technology and/or the internet was seen as a barrier, with midwives (n=32, 35.96%) expressing these concerns to a significantly greater extent than women (n=138, 17.90%) and partners (n=10, 10.42%). Support was identified as a benefit and a barrier. Regarding the former, mental health care support, whether that be in contrast or in addition to the current standard of care, was regarded a benefit by respondents, with a significantly higher proportion of midwives (n=36, 40.45%) perceiving this as a benefit relative to both women (n=139, 17.08%) and partners (n=18, 19.35%). However, just under 3% (n=25) of all respondents highlighted issues surrounding the potential quality of the support offered by the tool, as well as there needing to be appropriate follow-up care, particularly given the lack of current support and available services.

Overall, almost 50% (n=483) of respondents identified environmental opportunities, such as increased convenience, flexibility, and being able to complete the assessment in a comfortable environment, as benefits. Relative to midwives (n=27, 30.34%), a significantly higher proportion of women (n=417, 51.23%) regarded environmental opportunities as benefits. On the other hand, approximately 20% (n=198) of all respondents saw not having time or a comfortable and private environment where they felt safe as barriers to using the tool. Finally, funding and implementation were seen as both a benefit and a barrier. Around 2% (n=23) of all respondents identified the use of a digital mental health assessment as being cost-effective and resulting in reduced pressure on the NHS. On the contrary, issues surrounding costs to develop and implement the tool, including its integration into the health care system, and problems associated with bureaucracy and infrastructure in the NHS, were identified as barriers by almost 8% (n=76) of all respondents.

Automatic Motivation

Positive affect was identified as a benefit to using a digital mental health assessment. In particular, 20.48% (n=204) of all respondents saw the tool being less intimidating and intrusive, as well as less stressful to use, resulting in reduced anxiety relative to in-person care. On the contrary, negative affect was seen as a barrier, with issues surrounding the tool creating more distress and a general fear of the results being identified by approximately 3% (n=27) of all respondents.

Reflective Motivation

Control of one’s own mental health was reported as both a benefit and a barrier. Feeling empowered and being in charge of one’s own mental health were regarded as benefits by 7.33% (n=73) of all respondents. On the other hand, 6.90% (n=66) of all respondents also perceived having to take the initiative to prioritize one’s mental health as a barrier. Overall, approximately 4% (n=34) of all respondents saw it being easier to manipulate answers in order to obtain a certain outcome, as well as general issues surrounding dishonesty, as barriers. However, almost 6% (n=55) of all respondents also saw the potential to be more honest regarding mental health symptoms as a benefit.

In addition, reassurance, including feeling cared for and heard, was also mentioned as a benefit by almost 5% (n=45) of respondents, while issues surrounding the tool being impersonal were identified as barriers by around 13% (n=126) of respondents. Beliefs about the tool were seen as both benefits and barriers to its use. For instance, beliefs regarding the tool being trustworthy and evidence-based, as well as having the potential to revolutionize mental health care provision, were seen as benefits, with a significantly higher proportion of midwives (n=5, 5.62%) identifying this in comparison to the women (n=5, 0.61%) and partner (0%) groups. On the contrary, skepticism regarding the credibility of the tool and beliefs associated with a digital tool being inappropriate to diagnose mental health conditions were identified as barriers, with partners (n=10, 10.42%) expressing this concern to a significantly greater extent than women (n=35, 4.54%) and midwives (n=1, 1.12%). Finally, a general reluctance to using the tool, including being resistant to change, was seen as a barrier, with a significantly higher proportion of partners (n=7, 7.29%) identifying this relative to the women (n=14, 1.82%).

Discussion

Principal Findings

The ultimate goal of this research was to provide proof-of-concept support for the development and implementation of a digital mental health assessment to aid in the identification and triaging of perinatal mental health concerns in the United Kingdom. First, we explored the current state of perinatal mental health care provision, as well as in the context of the evolving COVID-19 pandemic, relative to NICE guidelines [1]. Second, we evaluated women’s, partners’, and midwives’ attitudes toward using a digital mental health assessment to screen, diagnose, and triage perinatal mental health concerns.

The vast majority of women and one-third of partners expressed having experienced poor mental health during the perinatal period, with the most typical conditions encountered by midwives throughout their careers being depression, anxiety, and PTSD. Critically, and in line with previous findings, rates of diagnoses were low relative to the number of individuals reporting concerns (eg, [29-31]). Perhaps more alarmingly, however, was the fact that the majority of respondents had referred themselves to an HCP, rather than being identified by a midwife as requiring further evaluation. Importantly, it has been shown that midwives may struggle to recognize perinatal mental health difficulties and are less inclined to refer women to a mental health specialist for further evaluation relative to other maternity HCPs [32]. This may be due, in part, to the low uptake of validated mental health screening tools in maternity care settings. Despite NICE guidelines [1] recommending the use of the Whooley Questions [33] and the 2-item Generalized Anxiety Disorder scale (GAD-2) [34] to screen for perinatal mental health problems, 60% of midwives in this study reported *not* using standardized questions or questionnaires.

Mental health care support following a miscarriage or termination or difficult birth was also largely overlooked, particularly when it came to partners. In fact, the vast majority of partners were not provided with any information or support throughout the perinatal period. This is in line with past research suggesting that partners often feel excluded in favor of a more women-centric approach to perinatal care [35-37]. Importantly, despite maternity care settings being women centric, the routine assessment of women's past or family history of mental health difficulties was poor and varied considerably across midwives. This presents a significant issue as the majority of women in this study expressed having had the condition *prior* to becoming pregnant. Indeed, women with a past history of mental disorders are often at risk of relapse and may need support with re‑emerging symptoms precipitated by pregnancy or postpartum [38-40]. NICE guidelines [1] recommend that *all* women should be routinely asked about past or present severe mental health conditions and treatment, as well as the presence of severe mental health in first-degree relatives. However, our findings indicate that, overall, these recommendations are not being followed in maternity care settings in the United Kingdom. This may be due to the many tasks that are normally performed during appointments, lack of training, discontinuity of care, and time constraints [41].

Critically, the COVID-19 pandemic has put additional pressure on midwives and has had a devastating impact on the mental health of patients. Over 60% of midwives reported having seen an increase in the number of women experiencing perinatal mental health difficulties. This was mirrored by the women, with 64% reporting poorer mental health since the outbreak. In contrast with these alarming figures, less than one-quarter of women had been asked about the effects of the pandemic on their mental health, and the majority of midwives reported there being no change in the extent to which mental health had been prioritized throughout the pandemic. However, this was closely followed by there being a greater emphasis on inquiring about mental health, highlighting inconsistencies in care provision across maternity care settings. Notably, midwives reported there being a general lack of guidance on how to best support women's perinatal mental health throughout the pandemic. Notwithstanding the recommendations by the Royal College of Obstetricians and Gynaecologists [13] on the use of remote means throughout the pandemic, only half of midwives reported using remote tools to support the needs of women who may be at risk of perinatal mental health difficulties, with less than one-third of women having discussed their mental health concerns with a midwife or HCP remotely.

Despite the low use of remote means to support perinatal mental health difficulties throughout the pandemic, there was a strong interest in using a digital mental health assessment to screen, diagnose, and triage perinatal mental health concerns, particularly among women and midwives. The digital assessment was seen to be well placed within maternity care settings, with an in-person only or a blended approach (ie, a combination of in-person and remote support) being preferred by women and partners in the event of further care being advised. The results of our COM-B analysis highlighted important implications for the development and implementation of a digital mental health assessment. Physical opportunities, such as increased convenience and flexibility relative to the current standard of care, were seen as some of the key benefits to using a perinatal digital mental health assessment, particularly among women. This result is in agreement with previous findings reported by a systematic literature review focused on digital health for perinatal care, which highlighted that patients saw the convenience of receiving care in their homely ambience as a benefit of digital interventions, with satisfaction rates varying between 86% and 95% in digital mental health studies and 90% in electronically monitored induced home births [42].

Increased accessibility to mental health care provision, including the potential for improved parity of care, was also reported as a benefit by the participants of our study. Importantly, accessibility was also regarded as a potential barrier to using the tool, especially among midwives, who expressed concerns regarding accessing care in remote or rural locations with poor internet connectivity, as well as issues surrounding the inclusion of individuals from a low socioeconomic background. Our findings on the theme of accessibility are in line with previous research on attitudes of women and health care professionals toward broadly defined mobile health interventions during pregnancy [43]. Mobile and digital technologies were perceived as a useful tool to increase access to care for at-risk women who could not attend perinatal clinics, but also as a potential source of digital exclusion among those who are socioeconomically disadvantaged.

Consequently, the perinatal digital mental health assessment should be designed to leverage perceived physical opportunity benefits while mitigating barriers. For instance, flexibility in terms of assessment completion and the option to complete the assessment in the comfort of one's home may increase engagement. Similarly, interoperability with various devices, the possibility to store assessment responses in the absence of good connectivity, or the possibility to conduct the assessment via text messaging may increase accessibility in hard-to-reach groups. Text messaging has often proved to be a suitable tool for the delivery of mental health services to rural and remote communities, and future research in this field should focus on improving the predictive models and computational linguistic tools that underlie the diagnostic and therapeutic value of text-based psychological services [44].

Positive affect (automatic motivation component) was also readily recognized as a benefit. Women, partners, and midwives highlighted positive emotions related to using a digital assessment, such as reduced anxiety and stress, and an increased sense of privacy. On the other hand, the perception of the assessment being impersonal (reflective motivation component) was commonly highlighted as a barrier. Thus, a personalized digital journey and the possibility of a follow-up assessment by an HCP may maintain the positive feelings of privacy while addressing the perceived lack of in-person care. Lastly, barriers were identified in relation to cognitive skills (psychological capability component), such as comprehension difficulties, poor self-reflection, and difficulties in expressing and acknowledging mental health issues. This was particularly concerning for midwives.

Therefore, it is crucial to design a tool that can cater to individuals who may find it difficult to complete a digital mental health assessment, such as those with poor comprehension skills, learning difficulties, or mental health symptoms that impair communication. This point was also raised in the research priority setting exercise coordinated by the James Lind Alliance Priority Setting Partnership for the “Digital Technology for Mental Health: Asking the Right Questions” project [44. The uncertainty around how mental health conditions can affect engagement with digital technologies was identified by patients, carers, and mental health care providers as a research question that must be addressed. Adopting a co-designing approach to development, where all potential users (eg, women and their families, midwives, mental health specialists) and stakeholders are regarded as collaborators, could help overcome barriers related to cognitive skills by ensuring that the product meets users’ needs and preferences whilst being clinically valid and feasible to deliver.

Taken together, the findings from this study highlight the urgent need to improve and standardize perinatal mental health care provision in the United Kingdom. While time is limited in maternity care settings, where midwives may feel pressured to prioritize physical over mental health, our study provides proof-of-concept support for the use of a digital mental health assessment tool as an innovative time- and cost-effective solution to the identification and treatment of perinatal mental health concerns. Digital technologies have the potential to support midwives in the recognition of perinatal mental health difficulties and are highly scalable. This means that support can be provided to the *family unit* as a whole, resulting in a comprehensive and scalable approach to the provision of mental health care throughout an often challenging time for women and their families.

Strengths and Limitations

The surveys were comprehensive and were carefully designed with input from an experienced perinatal psychiatrist and a practicing midwife. Furthermore, the use of mixed methods allowed for the revelation of information that would have not been obtained via quantitative research only. Additionally, the use of COM-B, a comprehensive supratheory model, helped develop a map of the broad landscape of behavior determinants that can be considered for the implementation of a digital mental health assessment. Patients’ and HCPs’ behaviors are often considered separately in behavior change literature. However, this introduces additional levels of complexity and the need for different models to understand a single intervention. The COM-B model, along with our choice of exploring both users’ and midwives’ views, offers a single comprehensive framework incorporating context and stakeholders that can be used to inform the design of future digital perinatal mental health assessment interventions.

Critically, the majority of the women and partners who participated in this study were highly educated and had an above average socioeconomic status, meaning that the findings from this study may not be generalizable to the broader UK population. Furthermore, individuals with mental health concerns and/or negative experiences with mental health care provision may have been more receptive to the recruitment materials and more likely to enroll in the study. As such, there is likely to be a recruitment bias. In addition, women and midwives were recruited separately and, as a consequence, associations between care provision and care experience should be interpreted with caution.

Acknowledgments

This study was funded by the Stanley Medical Research Institute (grant 07R-1888). We are most grateful to all the women, partners, and midwives for their participation. We are also grateful to perinatal psychiatrist Dr Rebecca Moore and practicing midwife Yariza Martin-Key for their valuable input.

Conflicts of Interest

SB is a director of Psynova Neurotech Ltd and Psyomics Ltd. NAM-K, TSS, and SB have financial interests in Psyomics Ltd.

Multimedia Appendix 1
Women and partners: sociodemographic and perinatal health characteristics, mental health provision, symptoms, and diagnosis, and COVID-19.
[DOCX File , 28 KB - jmir_v23i6e27132_app1.docx]

Multimedia Appendix 2
Midwives: sociodemographic characteristics, mental health provision, and COVID-19.
[DOCX File , 22 KB - jmir_v23i6e27132_app2.docx]

Multimedia Appendix 3
Benefits and barriers listed by codes and grouped by theme, COM-B component, and framework.
[DOCX File , 29 KB - jmir_v23i6e27132_app3.docx]

References

1. Antenatal and postnatal mental health: Clinical management and service guidelines. National Institute for Health and Care Excellence. 2014 Dec. URL: https://www.nice.org.uk/guidance/cg192 [accessed 2020-12-18]

2. Vigod SN, Villegas L, Dennis C, Ross LE. Prevalence and risk factors for postpartum depression among women with preterm and low-birth-weight infants: a systematic review. BJOG 2010 Apr;117(5):540-550. [doi: 10.1111/j.1471-0528.2009.02493.x] [Medline: 20121831]
3. Alipour Z, Lamyian M, Hajizadeh E. Anxiety and fear of childbirth as predictors of postnatal depression in nulliparous women. Women Birth 2012 Sep;25(3):e37-e43. [doi: 10.1016/j.wombi.2011.09.002] [Medline: 21959041]
4. Jones I, Chandra PS, Dazzan P, Howard LM. Bipolar disorder, affective psychosis, and schizophrenia in pregnancy and the post-partum period. Lancet 2014 Nov 15;384(9956):1789-1799. [doi: 10.1016/S0140-6736(14)61278-2] [Medline: 25455249]
5. Paulson JF, Bazemore SD. Prenatal and postpartum depression in fathers and its association with maternal depression: a meta-analysis. JAMA 2010 May 19;303(19):1961-1969. [doi: 10.1001/jama.2010.605] [Medline: 20483973]
6. Leach LS, Poyser C, Cooklin AR, Giallo R. Prevalence and course of anxiety disorders (and symptom levels) in men across the perinatal period: A systematic review. J Affect Disord 2016 Jan 15;190:675-686. [doi: 10.1016/j.jad.2015.09.063] [Medline: 26590515]
7. Bauer A, Parsonage M, Knapp M, Lemmi V, Adelaja B. The costs of perinatal mental health problems. LSE and Centre for Mental Health. 2014. URL: https://www.nwcscnsenate.nhs.uk/files/3914/7030/1256/Costs_of_perinatal_mh.pdf [accessed 2020-12-18]
8. Bayrampour H, McNeil DA, Benzies K, Salmon C, Gelb K, Tough S. A qualitative inquiry on pregnant women's preferences for mental health screening. BMC Pregnancy Childbirth 2017 Oct 03;17(1):339 [FREE Full text] [doi: 10.1186/s12884-017-1512-4] [Medline: 28974195]
9. Psaros C, Geller PA, Sciscione AC, Bonacquisti A. Screening practices for postpartum depression among various health care providers. J Reprod Med 2010;55(11-12):477-484. [Medline: 21291033]
10. Sambrook Smith M, Lawrence V, Sadler E, Easter A. Barriers to accessing mental health services for women with perinatal mental illness: systematic review and meta-synthesis of qualitative studies in the UK. BMJ Open 2019 Jan 24;9(1):e024803 [FREE Full text] [doi: 10.1136/bmjopen-2018-024803] [Medline: 30679296]
11. Lebel C, MacKinnon A, Bagshawe M, Tomfohr-Madsen L, Giesbrecht G. Elevated depression and anxiety symptoms among pregnant individuals during the COVID-19 pandemic. J Affect Disord 2020 Dec 01;277:5-13 [FREE Full text] [doi: 10.1016/j.jad.2020.07.126] [Medline: 32777604]
12. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
13. Guidance for antenatal and postnatal services in the evolving coronavirus (COVID-19) pandemic: information for health care professionals. Royal College of Obstetricians and Gynaecologist. 2020 Jul. URL: https://www.rcog.org.uk/globalassets/documents/guidelines/2020-07-10-guidance-for-antenatal-and-postnatal.pdf [accessed 2020-12-18]
14. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a Personal Smartphone Custom App to Assess the Patient Health Questionnaire-9 (PHQ-9) Depressive Symptoms in Patients With Major Depressive Disorder. JMIR Ment Health 2015;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]
15. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323 [FREE Full text] [doi: 10.1371/journal.pone.0084323] [Medline: 24465404]
16. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London, UK: Silverback Publishing; 2014.
17. Alexander KE, Brijnath B, Mazza D. Barriers and enablers to delivery of the Healthy Kids Check: an analysis informed by the Theoretical Domains Framework and COM-B model. Implement Sci 2014 May 23;9:60 [FREE Full text] [doi: 10.1186/1748-5908-9-60] [Medline: 24886520]
18. Han S, Middleton PF, Bubner TK, Crowther CA. Women's views on their diagnosis and management for borderline gestational diabetes mellitus. J Diabetes Res 2015;2015:209215 [FREE Full text] [doi: 10.1155/2015/209215] [Medline: 25785278]
19. Maidment DW, Ali YHK, Ferguson MA. Applying the COM-B Model to Assess the Usability of Smartphone-Connected Listening Devices in Adults with Hearing Loss. J Am Acad Audiol 2019 May;30(5):417-430. [doi: 10.3766/jaaa.18061] [Medline: 31044693]
20. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Barker F, Atkins L, de Lusignan S. Applying the COM-B behaviour model and behaviour change wheel to develop an intervention to improve hearing-aid use in adult auditory rehabilitation. Int J Audiol 2016 Jul;55 Suppl 3:S90-S98. [doi: 10.3109/14992027.2015.1120894] [Medline: 27420547]
23. Carney R, Bradshaw T, Yung AR. Physical health promotion for young people at ultra-high risk for psychosis: An application of the COM-B model and behaviour-change wheel. Int J Ment Health Nurs 2016 Dec;25(6):536-545 [FREE Full text] [doi: 10.1111/inm.12243] [Medline: 27432534]

24. Fulton EA, Brown KE, Kwah KL, Wild S. StopApp: Using the Behaviour Change Wheel to Develop an App to Increase Uptake and Attendance at NHS Stop Smoking Services. Healthcare (Basel) 2016 Jun 08;4(2) [FREE Full text] [doi: 10.3390/healthcare4020031] [Medline: 27417619]
25. McDonagh LK, Saunders JM, Cassell J, Curtis T, Bastaki H, Hartney T, et al. Application of the COM-B model to barriers and facilitators to chlamydia testing in general practice for young people and primary care practitioners: a systematic review. Implement Sci 2018 Oct 22;13(1):130 [FREE Full text] [doi: 10.1186/s13012-018-0821-y] [Medline: 30348165]
26. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. Nat Hum Behav 2020 May;4(5):451-459. [doi: 10.1038/s41562-020-0887-9] [Medline: 32377018]
27. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
28. 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]
29. Hewitt CE, Gilbody SM. Is it clinically and cost effective to screen for postnatal depression: a systematic review of controlled clinical trials and economic evidence. BJOG 2009 Jul;116(8):1019-1027. [doi: 10.1111/j.1471-0528.2009.02148.x] [Medline: 19438499]
30. Rafferty J, Mattson G, Earls MF, Yogman MW, COMMITTEE ON PSYCHOSOCIAL ASPECTS OF CHILD AND FAMILY HEALTH. Incorporating Recognition and Management of Perinatal Depression Into Pediatric Practice. Pediatrics 2019 Jan;143(1) [FREE Full text] [doi: 10.1542/peds.2018-3260] [Medline: 30559118]
31. Schwab W, Marth C, Bergant AM. Post-traumatic Stress Disorder Post Partum: The Impact of Birth on the Prevalence of Post-traumatic Stress Disorder (PTSD) in Multiparous Women. Geburtshilfe Frauenheilkd 2012 Jan;72(1):56-63 [FREE Full text] [doi: 10.1055/s-0031-1280408] [Medline: 25253905]
32. Jones CJ, Creedy DK, Gamble JA. Australian midwives' knowledge of antenatal and postpartum depression: a national survey. J Midwifery Womens Health 2011;56(4):353-361. [doi: 10.1111/j.1542-2011.2011.00039.x] [Medline: 21733106]
33. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. J Gen Intern Med 1997 Jul;12(7):439-445 [FREE Full text] [doi: 10.1046/j.1525-1497.1997.00076.x] [Medline: 9229283]
34. Gervais C, de Montigny F, Lacharité C, St-Arneault K. Where fathers fit in Quebec’s perinatal health care services system and what they need. Psychology of Men & Masculinity 2016 Apr;17(2):126-136. [doi: 10.1037/a0039558]
35. Lever Taylor B, Billings J, Morant N, Johnson S. How do women's partners view perinatal mental health services? A qualitative meta-synthesis. Clin Psychol Psychother 2018 Jan;25(1):112-129. [doi: 10.1002/cpp.2133] [Medline: 28857366]
36. Steen M, Downe S, Bamford N, Edozien L. Not-patient and not-visitor: a metasynthesis fathers' encounters with pregnancy, birth and maternity care. Midwifery 2012 Aug;28(4):362-371. [doi: 10.1016/j.midw.2011.06.009] [Medline: 21820778]
37. Stevens AWMM, Goossens PJJ, Knoppert-van der Klein EAM, Draisma S, Honig A, Kupka RW. Risk of recurrence of mood disorders during pregnancy and the impact of medication: A systematic review. J Affect Disord 2019 Apr 15;249:96-103. [doi: 10.1016/j.jad.2019.02.018] [Medline: 30769297]
38. MacLeod Hall C, Molyneaux E, Gordon H, Trevillion K, Moran P, Howard LM. The association between a history of self-harm and mental disorders in pregnancy. J Affect Disord 2019 Nov 01;258:159-162 [FREE Full text] [doi: 10.1016/j.jad.2019.06.062] [Medline: 31415929]
39. Micali N, Treasure J, Simonoff E. Eating disorders symptoms in pregnancy: a longitudinal study of women with recent and past eating disorders and obesity. J Psychosom Res 2007 Sep;63(3):297-303. [doi: 10.1016/j.jpsychores.2007.05.003] [Medline: 17719368]
40. Williams CJ, Turner KM, Burns A, Evans J, Bennert K. Midwives and women's views on using UK recommended depression case finding questions in antenatal care. Midwifery 2016 Apr;35:39-46. [doi: 10.1016/j.midw.2016.01.015] [Medline: 27060399]
41. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the Next-Generation Perinatal Care: An Overview of the Literature. J Med Internet Res 2018 Jun 05;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
42. Willcox JC, van der Pligt P, Ball K, Wilkinson SA, Lappas M, McCarthy EA, et al. Views of Women and Health Professionals on mHealth Lifestyle Interventions in Pregnancy: A Qualitative Investigation. JMIR Mhealth Uhealth 2015 Oct 28;3(4):e99 [FREE Full text] [doi: 10.2196/mhealth.4869] [Medline: 26510886]
43. Dwyer A, de Almeida Neto A, Estival D, Li W, Lam-Cassettari C, Antoniou M. Suitability of Text-Based Communications for the Delivery of Psychological Therapeutic Services to Rural and Remote Communities: Scoping Review. JMIR Ment Health 2021 Feb 24;8(2):e19478 [FREE Full text] [doi: 10.2196/19478] [Medline: 33625373]
44. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance Priority Setting Partnership. Lancet Psychiatry 2018 Oct;5(10):845-854. [doi: 10.1016/S2215-0366(18)30296-7] [Medline: 30170964]

Abbreviations

COM-B model: Capability, Opportunity, and Motivation Behavior model
GAD: Generalized Anxiety Disorder
HCP: health care professional
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
PHQ-9: Patient Health Questionnaire-9
PTSD: posttraumatic stress disorder

Edited by C Basch; submitted 14.01.21; peer-reviewed by M Gottschalk, H Pratomo; comments to author 03.03.21; revised version received 11.03.21; accepted 16.04.21; published 04.06.21.

Please cite as:
Martin-Key NA, Spadaro B, Schei TS, Bahn S
Proof-of-Concept Support for the Development and Implementation of a Digital Assessment for Perinatal Mental Health: Mixed Methods Study
J Med Internet Res 2021;23(6):e27132
URL: https://www.jmir.org/2021/6/e27132
doi:10.2196/27132
PMID:34033582

Original Paper

A Change Talk Model for Abstinence Based on Web-Based Anonymous Gambler Chat Meeting Data by Using an Automatic Change Talk Classifier: Development Study

Kenji Yokotani[1], DPhil

[1] Department of Clinical Psychology, Graduate School of Sciences and Technology for Innovation, Tokushima University, Tokushima-shi, Japan

Corresponding Author:
Kenji Yokotani, DPhil
Department of Clinical Psychology, Graduate School of Sciences and Technology for Innovation
Tokushima University
1-1, Minamijosanjimacho
Tokushima-shi, 770-0814
Japan
Phone: 81 88 656 7103
Fax: 81 88 656 7103
Email: yokotanikenji@tokushima-u.ac.jp

Abstract

Background: Change and sustain talks (negative and positive comments) on gambling have been relevant for determining gamblers' outcomes but they have not been used to clarify the abstinence process in anonymous gambler meetings.

Objective: The aim of this study was to develop a change talk model for abstinence based on data extracted from web-based anonymous gambler chat meetings by using an automatic change talk classifier.

Methods: This study used registry data from the internet. The author accessed web-based anonymous gambler chat meetings in Japan and sampled 1.63 million utterances (two-sentence texts) from 267 abstinent gamblers who have remained abstinent for at least three years and 1625 nonabstinent gamblers. The change talk classifier in this study automatically classified gamblers' utterances into change and sustain talks.

Results: Abstinent gamblers showed higher proportions of change talks and lower probability of sustain talks compared with nonabstinent gamblers. The change talk model for abstinence, involving change and sustain talks, classified abstinent and nonabstinent gamblers through the use of a support vector machine with a radial basis kernel function. The model also indicated individual evaluation scores for abstinence and the ideal proportion of change talks for all participants according to their previous utterances.

Conclusions: Abstinence likelihood among gamblers can be increased by providing personalized evaluation values and indicating the optimal proportion of change talks. Moreover, this may help to prevent severe mental, social, and financial problems caused by the gambling disorder.

(J Med Internet Res 2021;23(6):e24088) doi:10.2196/24088

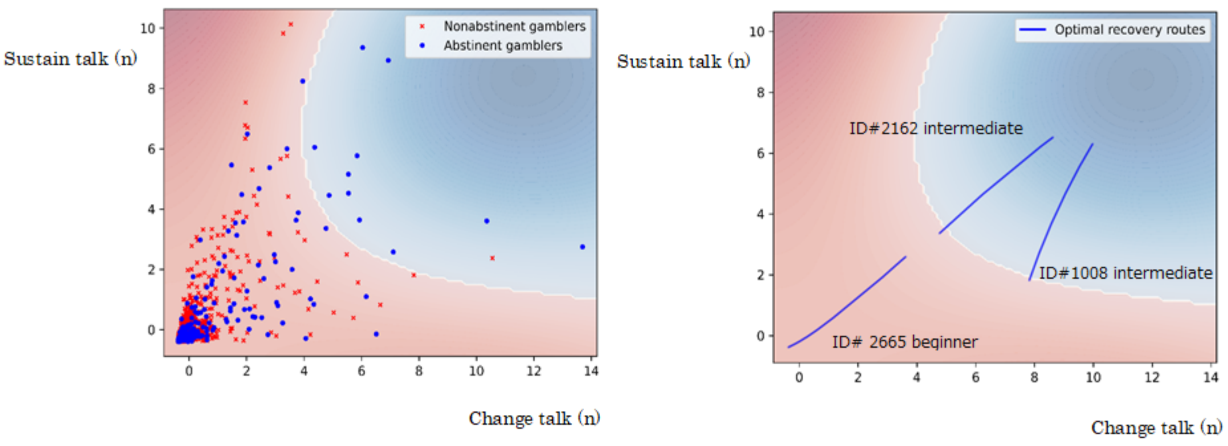
KEYWORDS

problem gambling; web-based anonymous gambler chat meetings; self-help group; change talk classifier; computerized text analysis; long-term data with dropout gamblers; recovery gradient; gradient descent method; gambling; addiction; abstinence

Introduction

Many individuals experience addictive disorders such as problematic drinking and gambling, and they require treatment to reduce the severity of their resulting mental and social problems [1,2]. One treatment option for such disorders is anonymous self-help group meetings, which have been proven to be as effective as standardized psychotherapies received in clinical settings [3]. Recent studies have shown that the abstinence process in anonymous self-help group meetings and standardized psychotherapies is similar [4], and the key factor in both treatments is to motivate participants to cease their addictive behaviors [5]. Studies have noted that individuals' motivation can be measured by classifying their utterances during therapy at individual web-based [6] and offline group settings [7]. Further, their utterances can be automatically classified using a machine-learned classifier [8]. Thus, the abstinence process in web-based anonymous self-help group meetings can be clarified through machine-learned classification of utterances. However, the web-based abstinence process is unclear [5] despite its usefulness in terms of ease-of-use [9] and few damages caused by prejudice [10]. To clarify the abstinence process, a change talk model for abstinence was developed in this study based on the utterances of anonymous web-based gamblers through the use of a machine-learned classifier. The talk model developed in this study can be used to visualize the abstinence process of a gambler and specify his/her individual abstinence likelihood as well as the type of utterance that will encourage or prevent him/her from depending on his/her previous utterances (Figure 1).

Figure 1. Change talk model for abstinence colored by evaluation values. Red areas indicate low abstinence likelihood areas (negative evaluation values), whereas blue areas indicate high abstinence likelihood areas (positive evaluation values). Horizontal and vertical lines represent standardized number of change and sustain talks, respectively. A: Scatter map of gamblers with and without current abstinent periods of at least three years. Blue circles and red Xs indicate abstinent and nonabstinent gamblers, respectively. B: Optimal recovery routes for 3 gamblers. Blue lines indicate 3 gamblers' continuous optimal ways to increase their likelihood of gambling abstinence responding to their minute status differences. To clarify the meanings of the number of change and sustain talks, the numbers were unstandardized in the following cases. The beginner gambler (ID# 2665) started at 1 change talk and 1 sustain talk with evaluation value –1.000 and finished at 1657.16 change talks and 351.31 sustain talks with evaluation value –0.588 after 500 trials. The evaluation value at the end remained negative, thereby indicating low likelihood of gambling abstinence. The best portion of change talks (number of change talks/number of change and sustain talks) for the beginner (ID# 2665) during web-based anonymous gambler chat meetings was regarded as 0.8254. The intermediate gambler (ID# 1008) started at 3396 change talks and 259 sustain talks with evaluation value –0.063 and finished at 4290.42 change talks and 787.67 sustain talks with evaluation value +2.815 after 500 trials. The evaluation value at the end turned positive, thereby indicating high likelihood of abstinence. The best portion of change talks for the intermediate gambler (ID# 1008) during web-based anonymous gambler chat meetings was regarded as 0.6285.



A

B

The theoretical framework used in this study is the self-perception theory [11]. This theory assumes that individuals are persuaded by their own utterances and will thus behave in accordance with it. Individuals with drug addiction who continuously stated that they would not use drugs during therapy showed lower drug reuse after 1 year compared with those who did not make such statements [12]. Similarly, individuals with alcohol addiction who continuously stated that they would drink during therapy indicated higher alcohol reuse after 1 year compared with those who did not make such statements [13]. These findings indicate that individuals' negative and positive comments on addictive behaviors were related to their improved and worse outcomes, respectively [14].

Based on these findings, motivational interviewing, a well-known method of motivating participants to cease addictive behaviors, classifies negative comments about addictive behaviors (eg, "Gambling is a waste of time") and positive comments about ceasing addictive behaviors (eg, "I can save money because I stopped gambling") together as "change talk" [14,15]. The results of a meta-analysis of motivational interviews showed that the proportion of participants' change talks during therapy was linked with improved outcomes for addictive behaviors [16]. Studies have shown that abstinent drug users showed more change talks during therapy compared with nonabstinent drug users [17]. Similarly, abstinent gamblers also showed more change talks during therapy compared with nonabstinent gamblers [18]. These findings indicate that gamblers' change talks are linked with their improved outcomes. Motivational interviewing also classifies positive comments about addictive behaviors (eg, "Gambling is the best game for me") and negative comments about ceasing addictive behaviors (eg, "Since I stopped gambling, I've become more irritable") together as "sustain talk" [14,19,20]. The results of a meta-analysis of motivational interviews showed that the number of participants' sustain talks during therapy was linked with negative outcomes for addictive behaviors [16]. Another meta-analysis including dissertations also confirmed the relation of sustain talks with negative outcomes [21]. One more study showed that gamblers' sustain talks were related with their negative outcomes [22]. These findings indicate that gamblers' sustain talks are linked with their negative outcomes.

The change and sustain talk classifications were primarily utilized for face-to-face motivational interviewing therapy settings [16,21-23]; however, recent studies have applied the classification to other standardized individual therapies [24], group therapies [7], and web-based therapies [6]. These findings indicate that the classifications can be applied to web-based anonymous gambler chat meetings. Moreover, individuals with 3-year continuous abstinent periods were considered more robust abstinent individuals than those who had 90-day abstinent periods [25,26], although several studies have considered individuals with 90-day abstinent periods as abstinent individuals [3]. Several gambling studies have used data related to gamblers with 1-year continuous abstinent periods rather than those with 90-day abstinent periods [27,28]. In this study, gamblers with at least 3-year continuous abstinent periods were regarded as abstinent gamblers.

Based on studies on change and sustain talk classifications [16,21-23] and the change talk classifier [8], the change talk model in this study was developed to differentiate abstinent and nonabstinent gamblers. Recent studies have shown that instead of evaluating change and sustain talks separately, both must be evaluated simultaneously [16,22,23]. Hence, our study model considers the number of change talks and sustain talks as x and y axes variables in a scatter plot, respectively (Figure 1). Furthermore, the proportion of change talks (number of change talks/number of change and sustain talks) was evaluated in this study because studies have indicated that the proportion of change talks was a better index of improved outcomes than the number of change talks [16,22,23]. To validate this study's findings with those of previous studies [16,22,23], this study hypothesized that the proportion of change talks among abstinent gamblers will be higher than that in nonabstinent gamblers in web-based anonymous gambler chat meetings.

Methods

This study used registry data from the internet.

Data Source

The author of this study accessed web-based anonymous gambler chat meeting data in Japan [29]. The meetings were conducted since September 2008 through web-based text chats with no pictures or sounds. A few offline meetings were conducted for a year in Osaka, Tokyo, or Sapporo. The web-based anonymous gambler chat meetings has no relationship with the Gamblers Anonymous group [28]. To post in the meetings, users had to be approved by the administrator in advance. Individuals who could participate in the meetings were those who were experiencing a gambling problem or their family members who were experiencing a gambling problem. The meetings were conducted by gamblers without any strict rules, and no outside experts were involved at all. The meetings were divided into those for users aiming for 1 week, 1 month, 3 months, 6 months, 1 year, 3 years, and more than 3 years of gambling abstinence, and users attended the meetings based on the number of days they had remained gambling-abstinent. However, these meetings had no strict rules, and a user aiming for 3-month gambling abstinence may attend a meeting for those aiming for 1-month gambling abstinence and vice versa. There were meetings for chatting as well as for providing daily reports and reports of memorial days (such as 100 abstinent days). Furthermore, these meetings were always open and free to join; therefore, the time and duration of participation varied widely among users (Table 1). Table 1 shows a comparison of the demographic variables, gambling history, gambling symptoms, participation forms, and characteristics of the utterances in web-based anonymous gambler chat meetings between abstinent and nonabstinent gamblers. In Table 1, abstinent gamblers refers to gamblers who have remained abstinent without relapse for at least three years. The utterance classifier in this study involves 6 clusters, but only participants' change talk and sustain talk clusters were used in this study. The data size of the demographic variables and gambling histories was smaller than that of the utterances. The proportion of change talks was the number of change talks per the total number of change and sustain talks. Some users did not have any change and sustain talks; thus, their proportions were not calculated. When a single meeting exceeded 5000 posts, it would end and a new meeting would be started. These meetings spanned multiple days or months and were rarely completed within a day.

Table 1. Comparison of the demographic variables, gambling history, gambling symptoms, participation forms, and characteristics of the utterances in web-based anonymous gambler chat meetings between abstinent and nonabstinent gamblers.

Characteristics	Abstinent gamblers (n=267), mean (SD)	Nonabstinent gamblers (n=1625), mean (SD)	*d*	*t (df)*	*P* value
Demographic variables					
Age (years)	35.000[a] (9.520)	35.811[b] (9.440)	–0.086	–0.896 (172.806)	.37
Proportion of males	0.845[c]	0.860[d]	N/A[e]	–0.565 (256.025)	.57
Gambling history					
Total amount of debt (million ¥)[f]	3.718[g] (6.602)	2.159[h] (3.727)	0.003	1.177 (27.320)	.25
Length of gambling (years)	15.301[i] (23.030)	11.843[j] (7.350)	0.202	1.272 (74.526)	.21
Gambling symptoms					
Number of symptoms (min:1, max:10)	2.981 (1.801)	2.661 (1.766)	0.179	2.696 (355.278)	.01
Gambling tolerance	0.408 (0.492)	0.371 (0.483)	0.076	1.146 (355.382)	.25
Unsuccessful control over gambling	0.730 (0.445)	0.733 (0.442)	–0.007	–0.104 (358.057)	.92
Preoccupied with gambling	0.667 (0.472)	0.542 (0.498)	0.257	3.969 (370.285)	<.001
Lies in gambling	0.311 (0.464)	0.198 (0.399)	0.261	3.750 (333.748)	<.001
Reliance on others to provide money	0.094 (0.292)	0.079 (0.270)	0.051	0.747 (345.207)	.46
Illegal acts for gambling	0.004 (0.061)	0.040 (0.196)	–0.250	–5.907 (1309.252)	<.001
Participation forms in web-based anonymous gambler chat meetings					
Participation length (days)	842.974 (904.471)	386.072 (616.142)	0.590	7.957 (307.817)	<.001
Interval length (days)	16.168 (63.090)	24.241 (104.712)	–0.093	–1.735 (543.061)	.08
Utterance characteristics in web-based anonymous gambler chat meetings					
Number of total utterances	2395.625 (5146.820)	613.065 (1797.483)	0.462	5.603 (276.750)	<.001
Number of change talks	410.007 (764.597)	115.869 (303.165)	0.506	6.206 (279.889)	<.001
Number of sustain talks	98.944 (192.787)	37.585 (97.324)	0.402	5.095 (288.661)	<.001
The proportion of change talks	0.742[k] (0.200)	0.681[l] (0.231)	0.280	4.477 (392.158)	<.001
Average probability of change talks	0.298 (0.155)	0.299 (0.160)	–0.001	–0.022 (365.505)	.98
Average probability of sustain talks	0.096 (0.062)	0.122 (0.076)	–0.369	–6.017 (406.651)	<.001
The proportion of change talks' probability	0.730 (0.150)	0.685 (0.167)	0.285	4.495 (382.728)	<.001

[a] n=129.

[b] n=771.

[c] n=193.

[d] n=1118.

[e] N/A: not applicable.

[f] ¥1 million=US $9103.

[g] n=26.

[h] n=182.

[i] n=73.

[j] n=427.

[k] n=266.

[l] n=1612.

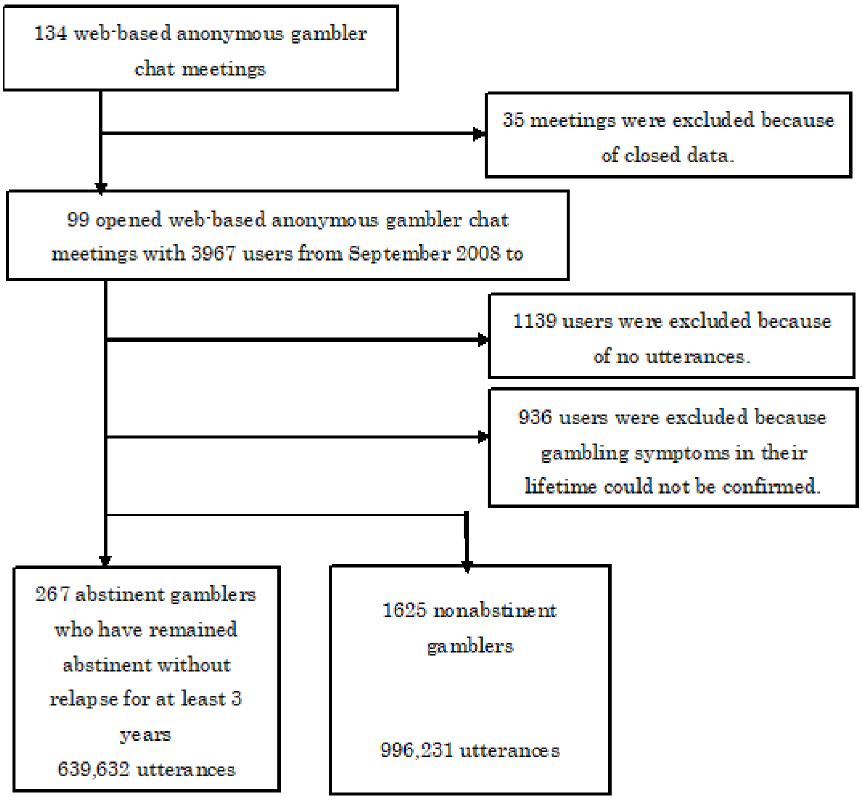
Sample Size

Research focusing on 3-year continuous abstinent periods is limited [27]; therefore, this study used 3-year-continuous abstinent smokers' studies [25] because smokers and gamblers share similar biological and environmental factors [30]. In the previous study [25], 13.8% of smokers achieved 3-year abstinence with the help of treatment whereas 86.2% did not. Based on the abovementioned findings [25], the allocation rate between abstinent and nonabstinent gamblers was set as 0.16. To show satisfactory power (0.95) with alpha (.05) for small effect size (*d*=0.2), 2736 participants were required.

Participants

A total of 134 web-based anonymous gambler chat meetings were web scraped on March 10, 2020 (Figure 2) [29]. Among them, 35 meetings included private data and were thus excluded. From the 99 remaining meetings, 3967 users were identified based on their anonymous names. They posted Japanese texts in the meetings from September 8, 2008 to March 10, 2020. Among them, 1139 gamblers were excluded because their posts included lesser than 3 words or included only advertisements with "http." Furthermore, 936 users were excluded because we could not confirm the gambling symptoms in their lifetime based on their posts (the evaluation of gambling symptoms is presented in the next section). The final participants were 1892 gamblers who experienced at least one gambling symptom in their lifetime. The number of study participants (n=1892) did not reach the ideal sample size (n=2736). To show satisfactory power (0.95) with alpha (.05) in the 1892 participants, the effect size must be over 0.24.

Figure 2. Data extraction of the participants.



Collection of Participants' Basic Data

Gambling Symptoms

Among the users' posts, texts were extracted based on keywords related to 10 gambling symptoms, that is, 1 gambling symptom related to illegal behaviors in gambling [31] and 9 symptoms of gambling disorder as mentioned in the fifth edition of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders [1]. For example, "lie" is a keyword used for examining the symptom of lies in gambling; therefore, all texts involving "lie" were extracted. Two raters who majored in clinical psychology as undergraduate students and received 2-hour training regarding gambling disorders from a Japanese clinical psychologist [32] independently read the texts and evaluated whether the user had experienced the aforementioned 10 symptoms in their lifetime. To check the validity of their evaluation, the psychologist also blindly evaluated randomly selected texts from 100 participants. Kappa coefficients confirmed that the psychologist and the 2 raters were primarily in "almost perfect or perfect agreement" [33] in their evaluation of the presence of gambling symptoms (0.835 and 0.860, Table S1 of Multimedia Appendix 1). Based on their assessments, participants with at least one gambling symptom in their lifetime were included as the final participants (Figure 2). Furthermore, the number of gambling symptoms as well as the gambling symptoms for which agreement between the psychologist and the 2 raters was above “substantial agreement” (0.60) [33] are also listed in Table S1 (Multimedia Appendix 1) as the basic data for the user.

Demographic Variables

The sex and age of the participants were estimated using the first 1000 words of the aforementioned gamblers because these words frequently involved self-introduction statements. One-third of these gamblers were determined using a Japanese application [34] (Table 1).

Gambling History and Debt

Among the gamblers’ words, text was extracted based on the keywords, including “years” and “history,” and the text was read by the 2 raters. If the text revealed a gambling history, the user’s years of gambling were recorded (Table 1). Similarly, the text was extracted based on keywords such as “yen,” “ten thousand,” “cash,” or “borrow.” If the text indicated a debt, the debt was listed (Table 1). Missing data were excluded.

Participation Forms in Web-Based Anonymous Gambler Chat Meetings

The length of the participation in the meetings was estimated based on the days of the initial and final posts in the meetings. Furthermore, participants’ posts in the meetings involving long paragraphs were divided into 2 sentence units. If sentences were lesser than 3 words or included “http,” they were removed. A two-sentence text will herein be referred to as an utterance. The average interval of utterances was also estimated based on users’ participation length and number of utterances.

Outcomes

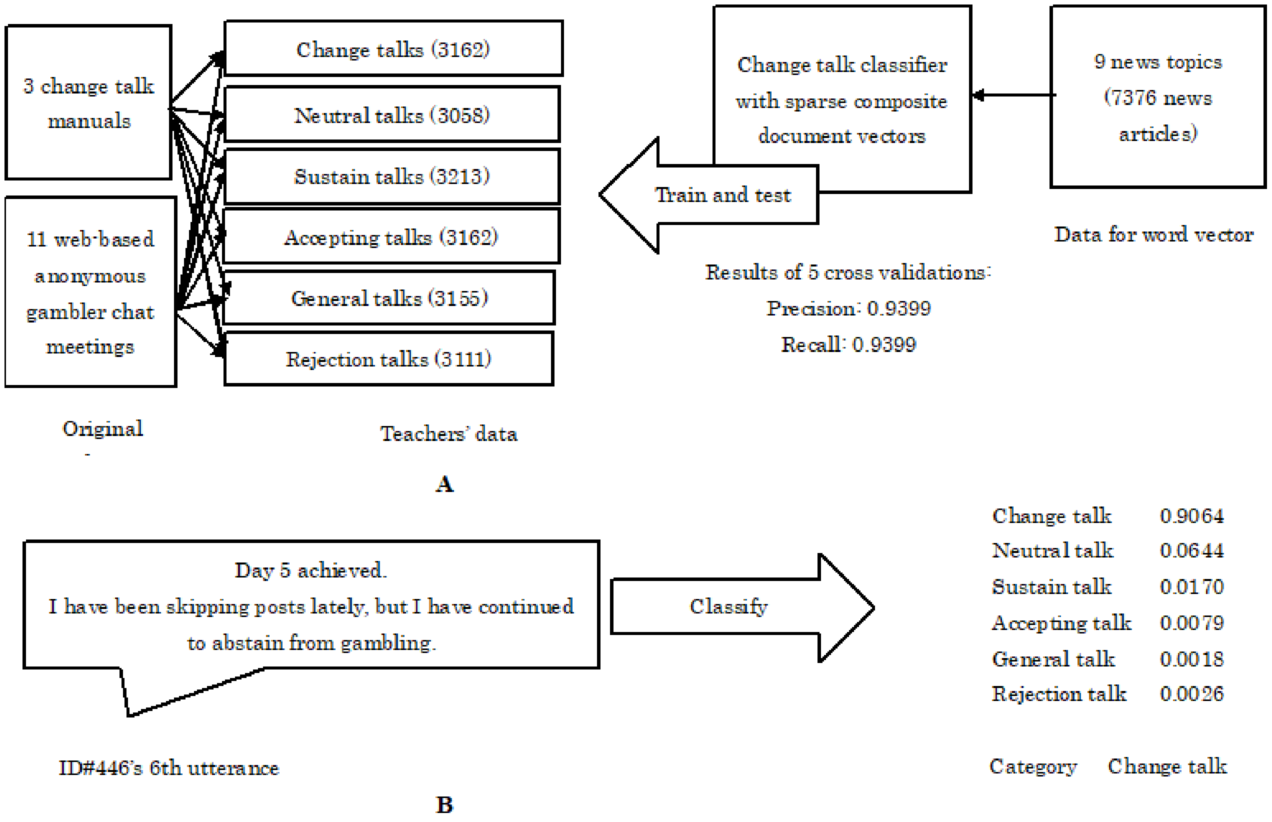
Users who registered as gamblers were provided with a personal counter that automatically counts the number of days wherein they ceased gambling. If they gamble, they report it to the administrator who resets the number of days on the counter. Those whose counters exceed 3 years are listed on a separate website and exemplified as models who have stopped gambling. According to the list on March 2020, 267 gamblers were regarded as abstinent gamblers without relapse for at least 3 years with 639,632 utterances; the other 1625 gamblers were regarded as nonabstinent gamblers with 996,231 utterances. These counters and lists were created by the administrator without the involvement of the author.

Implementation of the Change Talk Classifier

General Schema of the Change Talk Classifier

To create a change talk classifier, sparse composite document vectors were used [35]. A total of 7376 Japanese news articles were used to create a 200-dimensional word vector [36] (Figure 3A). To construct the document vector, 18,861 utterances were used with 6 clusters: clients’ change, neutral, and sustain talks, as well as therapists’ accepting, general, and rejection talks [15,19,20] (Figure 3A). Although each cluster has subcategories, the utterances were classified into clusters instead of subcategories. Furthermore, half of the utterances in each cluster was derived from 3 change talk manuals [15,19,20]; the other half came from the initial parts of 11 web-based anonymous gambler chat meetings [29].

Figure 3. Implementation of an automatic change talk classifier. A: Data sets for the change talk classifier; B: An example of a change talk classifier.



Creation of Teachers' Data

Comments in the 3 manuals focused on other addictive behaviors [15,19,20] and were converted into gambling topics, for example, the comment "I am going to stop smoking tomorrow" was changed to "I am going to stop gambling tomorrow." Additionally, several types of gambling are noted in Japan; therefore, the word "gambling" was used to replace 17 specific gambling terms such as "pachinko (Japanese pinball) game" and "horse racing." In addition, the Japanese language includes several types of second personal pronouns [37]; therefore, 3 typical personal pronouns were used, namely, "anata," "satosan," and the zero pronoun. Each comment in the manual was converted into 51 different gambling comments (17 gamble types × 3 pronoun types). Finally, 31,977 utterances related to 6 clusters were obtained from the manuals. Furthermore, 9519 utterances were identified during the initial parts of the 11 web-based anonymous gambler chat meetings. These utterances were encoded by a Japanese clinical psychologist with expertise in mail-based motivational interviewing [38]. Finally, 9519 utterances in the meetings related to the 6 clusters were obtained.

Preprocessing of Teachers' Data

Comments longer than 2 sentences were considered as an utterance. When the comment comprised a single sentence, the sentence was randomly combined with another comment from the same type of gambling (eg, "pachinko"), pronoun (eg, "anata"), and subcategory (eg, "ability in change talks") utterances. The following 2 points were included as constraints. First, one cluster from one source should have 1500 utterances to reflect the rate of comments within the subcategories. Second, if the number of utterances in a subcategory was lesser than 51 because of the first constraint, the utterances should continue until they exceed 51. Finally, we obtained 18,861 utterances tied with 6 clusters. During the training and testing of the change talk classifier in this study, the number of words in each utterance was fixed at 200 to remove the effects of this aspect on the classification. For utterances of more than 200 words, the exceeding words were removed. For utterances lesser than 200 words, the 2 sentences were repeated until they reached 200 words.

Performance of the Change Talk Classifier

Among the 18,861 utterances in 6 clusters (Figure 3A), 4/5 were used as training data and 1/5 was used as test data. The training and test sessions were repeated 5 times. The average scores of precision and recall during these 5 sessions were 0.9399 and 0.9399, respectively. Utterances in web-based anonymous gambler chat meetings were labeled based on the classifier used in this study. For example, the classifier showed the estimated probability of the following utterance: "Day 5 achieved. Although I have been skipping posts lately, I have continued to abstain from gambling" (Figure 3B). The estimated probability of change talk was 0.9064, which is the maximum probability among the 6 clusters. Hence, the utterance was labeled as change talk (Figure 3B). Similarly, 1,635,863 utterances in the meetings were classified into 6 clusters. The classifier in this study did not involve the list of abstinent gamblers; the classifier was blind to the outcome.

Independent Variables

Change Talk

Based on the change talk classifier, the number of change talks among 1892 gamblers was evaluated (Table 1). The estimated probability that the classifier uses for classification was utilized (Figure 3B) and showed the average probability of the change talk (Table 1).

Sustain Talk

Similarly, the number of sustain talks was evaluated (Table 1). The estimated probability for the classification was utilized (Figure 3B) and showed the average probability of the sustain talk (Table 1).

Proportion of Change Talks

Based on previous studies [16,22,23], the proportion of change talks (number of change talks/number of change and sustain talks) was evaluated in this study. Similarly, the proportion of change talk probabilities (average probability of change talks/average probabilities of change and sustain talks) was evaluated.

Statistical Analysis

Model Comparison

Our change talk model for abstinence involved the number of change and sustain talks as independent variables and 3-year abstinence as outcomes (Figure 1). The number of change and sustain talks was high; therefore, they were standardized in the model (Figure 1). To find the best identification function for abstinent and nonabstinent gamblers, a support vector machine (SVM) was compared with a linear and (nonlinear) radial basis kernel function (RBF). The cost of penalty and complexity of boundary area parameters were estimated using the grid search method [39] from the set of 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, and 10^{1}. The 3 cross-validation accuracies of SVM with RBF (0.8555, where cost of penalty and complexity of boundary area were 10^{1} and 10^{-1}, respectively) outperformed the accuracy of SVM with linear function (0.8516, where cost of penalty and complexity of boundary area were 10^{-3} and 10^{-3}, respectively). Hence, we used the SVM with RBF as the identification function for abstinent and nonabstinent gamblers. The SVM has an evaluation value that assesses the abstinence likelihood of gambling, with a negative score indicating low abstinence likelihood and a positive score indicating high abstinence likelihood (Figure 1).

Visualization of the Optimal Abstinence Process

To visualize the differences in the optimal abstinence processes of gamblers, the gradient descent method was used:

where x_i and y_i are the standardized number of change and sustain talks in trial i, respectively; f is the SVM identification function with RBF; and are partial differentials of f at the point x_i and y_i, respectively; and e is the learning rate set as 0.01. The maximum number of trials was set to 500.

Furthermore, the current x_i and y_i monotonically increase and the area of the partial differential is limited within the first quadrant. Moreover, when the score of f at trial i was better than that at trial $i+1$, the score of f at i was considered a local solution and the trials were ceased.

Software Used

For sample size estimate and power analysis, G*power 3.1.9.4 was used [40]. To determine users' age and sex through text, COTOHA [34] was used. For Japanese text analysis, MeCab [41] was used. The original Python codes [35] were used to create the document vector.

Ethical Considerations

This study was approved by the ethics committee of a National University in Japan (Reception number 222, ethical review on September 1, 2020). Furthermore, all procedures were conducted in accordance with the guidelines for studies involving human participants, the ethical standards of the institutional research committee, and the revised 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Results

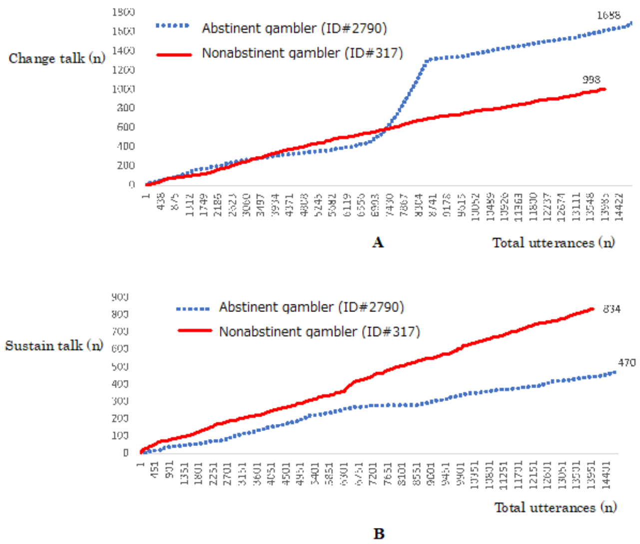
Comparison of the Basic Characteristics Between Abstinent and Nonabstinent Gamblers

The gamblers examined in this study were primarily males (1125/1311, 85.8%) and aged around 35 years (Table 1). Although their sex, age, total amount of debt, and history of gambling were not significantly different, abstinent gamblers who had remained abstinent without relapse for at least three years experienced significantly more gambling symptoms (P=.01) than nonabstinent gamblers (Table 1). Although their gambling symptoms such as gambling tolerance, unsuccessful control over gambling, and reliance on others to provide money were not significantly different, abstinent gamblers had significantly higher preoccupation (P<.001) with gambling and lies for gambling compared with nonabstinent gamblers. The nonabstinent gamblers were significantly more experienced in illegal acts of gambling (P<.001) than abstinent gamblers. These findings indicated that abstinent gamblers experienced more gambling symptoms compared with nonabstinent gamblers although the effect size was small (d=0.179-0.261).

Comparison of the Utterances in Web-Based Anonymous Gambler Chat Meetings Between Abstinent and Nonabstinent Gamblers

Abstinent gamblers participated significantly longer in the meetings and talked more than nonabstinent gamblers (Table 1, both P<.001, d=0.590 and d=0.462, respectively). The average probability of change talks was not significantly different between abstinent and nonabstinent gamblers; however, the average probability of sustain talks in nonabstinent gamblers was significantly higher than that in abstinent gamblers (Table 1, P<.001, d=–0.369). Hence, the proportion of change talks' probabilities (average probability of change talk/average probabilities of change and sustain talks) in abstinent gamblers was significantly higher than that in nonabstinent gamblers (Table 1, P<.001, d=0.285). Similarly, the proportion of change talks (number of change talks/number of change and sustain talks) in abstinent gamblers was significantly higher than that in nonabstinent gamblers (Table 1, P<.001, d=0.280). These findings indicate that abstinent gamblers showed higher proportion of change talks in the meetings compared with nonabstinent gamblers. To clarify the proportion of change talks between abstinent and nonabstinent gamblers, dynamic differences in the utterances between abstinent and nonabstinent gamblers were compared with similar number of utterances. Figure 4A shows the accumulated numbers of change talks between abstinent (ID# 2790) and nonabstinent (ID# 317) gamblers. Figure 4A indicates that abstinent gamblers produced more change talks than nonabstinent gamblers. Figure 4B shows the accumulated numbers of sustain talks between them. In contrast with that seen in change talks, nonabstinent gamblers produced more sustain talks than abstinent gamblers. Figure 4 shows that the utterances of abstinent and nonabstinent gamblers in the meetings showed dynamic differences.

Figure 4. Dynamic differences in the change and sustain talks between abstinent and nonabstinent gamblers in web-based anonymous gambler chat meetings. Blue and red lines in A and B indicate abstinent (ID# 2790) and nonabstinent (ID# 317) gamblers, respectively. A: The accumulated number of change talks during web-based anonymous gambler chat meetings. The vertical and horizontal lines indicate the number of change talks and total utterances, respectively. B: The accumulated number of sustain talks during web-based anonymous gambler chat meetings. The vertical and horizontal lines indicate the number of sustain talks and total utterances, respectively.



Development of the Change Talk Model for Abstinence

To visualize the dynamic differences between abstinent and nonabstinent gamblers, the change talk model for abstinence, involving the number of change and sustain talks as independent variables on x and y axes, respectively, classified 267 abstinent gamblers and 1625 nonabstinent gamblers (Figure 1A). Before classification, the correlations among the number of change talks, the number of sustain talks, and gambling abstinence were checked. The number of change and sustain talks were positively correlated with gambling abstinence (r=0.247, r=0.182, both were P<.001, n=1892). Moreover, the number of change talks was positively correlated with the number of sustain talks (r=0.676, P<.001, n=1892). These correlations indicated the positive relation of these talks with gambling abstinence.

Figure 1A shows abstinent and nonabstinent gamblers scattered on a map with the number of change and sustain talks on x and y axes, respectively. The abstinence likelihoods of gambling were also color-coded based on the SVM with RBF (accuracy was 0.8555, where cost of penalty and complexity of boundary area were 10^1 and 10^{-1}, respectively). The strong red color indicates low abstinence likelihood (negative evaluation scores) whereas the strong blue color indicates high abstinence likelihood (positive evaluation scores). Figure 1A shows that participants who increased the number of change and sustain talks were likely to move from the red zone to the blue zone, which indicated that they increased their abstinence likelihood of gambling.

Clinical Use of the Change Talk Model for Abstinence

To clarify the clinical use of our model, 1 beginner and 2 intermediate gamblers were plotted, and the individual evaluation values for abstinence as well as their ideal proportion of change talks were shown through the gradient descent method (Figure 1B). To clarify the meanings of the number of change and sustain talks, the numbers were unstandardized. The beginner gambler (ID# 2665) started at 1 change talk and 1 sustain talk with evaluation value –1.000 and finished at 1657.16 change talks and 351.31 sustain talks with an evaluation value of –0.588 after 500 trials. The beginner gambler's (ID# 2665) evaluation value at the end remained negative, thereby indicating low abstinence likelihood. The ideal portion of change talks (number of change talks/number of change and sustain talks) for this gamble during the web-based anonymous gambler chat meetings was regarded as 0.8254. One intermediate gambler (ID# 2162) started at 2130 change talks and 441 sustain talks with evaluation value –0.1167 and finished at 3731.94 change talks and 813.04 sustain talks with evaluation value 2.507 after 500 trials. His evaluation value became positive at the end, thereby indicating high abstinence likelihood. The best portion of change talks for him during the meeting was considered 0.8115. The other intermediate gambler (ID# 1008) started at 3396 change talks and 259 sustain talks with evaluation value –0.063 and finished at 4290.42 change talks and 787.67 sustain talks with evaluation value +2.815 after 500 trials. This gambler's evaluation value turned positive at the end, thereby indicating high abstinence likelihood. The best portion of change talks for this gambler during the meetings was considered 0.6285. These findings indicate that the evaluation values are personalized feedback for gamblers. Furthermore, high proportions of change talks (over 0.80) were required for most beginner gamblers, whereas the best proportions of change talks were different among intermediate gamblers.

Discussion

Principal Results

The change talk classifier used in this study indicated 93% precision and analyzed 1.63 million utterances in web-based anonymous gambler chat meetings (Figure 3), thus leading to the development of the change talk model for gambling abstinence. The abstinent map with evaluation values showed the abstinence likelihoods of each gambler (Figure 1). Based on suggestions from a previous study on anonymous group meetings [4,5], this study confirmed that the abstinence process in web-based anonymous gambler chat meetings was similar to the process in other standardized therapies. Consistent with the results of previous studies regarding face-to-face motivational interviewing for gamblers [16,18,22,23], the proportion of change talks was also linked with improved outcome. In particular, the proportions at the initial stages were positively linked with 3-year continued gambling abstinence. This study advances previous findings regarding language classifiers in mental health fields [42,43] in 2 ways. First, the classifier used in previous studies classified the onset of mental disorders, whereas the classifier in this study classified recovery from a mental disorder. Second, the classifiers used in previous studies served as a screening tool for the early detection of mental disorders, whereas the classifier in this study is a therapeutic tool for identifying the recovery process. The 2D map with the abstinence likelihood slope used in this study was helpful for showing personalized evaluation values and ideal proportions of change talks, responding to every gambler's utterance in the meetings.

This study involves 2 new methodologies: a change talk classifier and long-term data with dropout gamblers. First, this study used machine-learning methods and automatically classified gamblers' utterances (Figure 3) [35]. Studies have utilized human expert resources to classify gamblers' utterances [12,15,18-20,44-47], with limited data size (around 10^3) and generalization of their findings. The change talk classifier used in this study increased data size (over 10^6) similar to recent mental health research [8,42,43] and was applicable to other therapy processes focusing on change talks such as group and individual therapies for healthy dieting and prevention of substance abuse [16,23]. The utilization of the change talk classifier for these therapies can save human resources about change talk classification [12,19,20] and broaden the applicability of change talk classifications or the core scheme of motivational interviewing [14-16] for addictive disorders [7,21].

This study also involved long-term data (maximum participation length of 3872 days) with both dropout and ongoing data sets (Figure 2, Table 1). Inclusion of these data sets enabled extensive and detailed slope maps of abstinence likelihoods of gambling. Studies have primarily utilized the initial and final sessions' scores of gamblers who completed the treatments. However, considering their high dropout rates [48,49], these data sets could bias gamblers' characteristics. Moreover, the data set involving only the first and last sessions does not show the processes in between, and the model based on these data sets does not reveal the abstinence likelihood corresponding to the therapy processes [50]. Long-term data with dropout and ongoing data sets can be helpful for estimating changes during therapy, particularly for diseases that require extensive treatment periods [25-28].

Limitations

This study has 5 limitations. First, the number of female participants in the study sample was limited. The abstinence process differs between female and male participants [51]; therefore, future studies must involve more female participants. Second, the current measurement of gamblers' symptoms was based on their utterances in web-based anonymous gambler chat meetings. Hence, our measurement may underestimate several gambling symptoms because the measurement could not assess these symptoms that gamblers did not reveal in the meetings. To compensate for these limitations, questionnaires [32] or standardized interview [31] data are required in future studies. Third, this study identified gamblers through their anonymous names, but it is possible that 2 different gamblers used the same anonymous name or that one gambler used multiple anonymous names. Their identities could not be checked because of the web-based anonymity settings. Another data set such as physical meetings [24,49] would be helpful for validating the findings of this study. Fourth, the small number of long-term participants is also one of the study limitations. This study model shows the highest likelihood area for gambling abstinence (Figure 1A, top right most blue area); this implies that individuals in this area will not benefit from participating in web-based anonymous gambler chat meetings. This point must be re-examined by collecting more data regarding long-term participants. Fifth, the internal validity of this study was limited because abstinent and nonabstinent gamblers freely interacted and the proportion of their change talks was not manipulated. To increase the internal validity, future studies must consider an experimental design.

Conclusion

Despite these limitations, this study developed the change talk model for gambling abstinence in web-based anonymous gambler chat meetings with the change talk classifier. The change talk classifier used in this study increased the data size of psychotherapies [6,12,13,18,44-47] and broadened the applicability of language analysis on social media for mental health services [42,43]. The change talk model for abstinence used in this study can provide personalized evaluation scores for abstinence and an ideal proportion of change talks to the gamblers responding to every utterance during the meetings. In clinical settings, using this model, participants can determine their abstinence likelihood because the evaluation values are determined based on their previous utterances before they participate in the meetings (Figure 1B). The ideal proportion of change talks can be estimated using this model (Figure 1B) and they can be motivated to comment at the meetings based on that proportion. After the meeting, gamblers can also observe changes in their evaluation values during the meeting, thus allowing them to objectively evaluate their individual progress of each meeting (Figure 1B). These factors are considered to provide personalized feedback for participants [14,17] and can likely lead to improved outcomes [38]. Providing personalized evaluation values and the optimal proportion of change talks, which were also validated in millions of data sets, can increase the abstinence likelihood of gamblers. This may help to prevent severe mental, social, and financial problems caused by the gambling disorder [1-3].

Acknowledgments

This study was funded by a grant from KAKENHI (19K11206). The author appreciates the insightful comments on the early draft from Prof. Sato, Yutaka.

Conflicts of Interest

The author was funded by a grant from KAKENHI (19K11206).

Multimedia Appendix 1
Supplementary data.
[DOCX File , 14 KB - jmir_v23i6e24088_app1.docx]

References

1. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Washington DC: American Psychiatric Pub Inc; 2013.
2. McLellan A, Lewis D, O'Brien CP, Kleber H. Drug dependence, a chronic medical illness: implications for treatment, insurance, and outcomes evaluation. JAMA 2000 Oct 04;284(13):1689-1695. [doi: 10.1001/jama.284.13.1689] [Medline: 11015800]
3. Project MATCH Research Group. Matching alcoholism treatments to client heterogeneity: Project MATCH three-year drinking outcomes. Alcohol Clin Exp Res 1998 Sep;22(6):1300-1311. [doi: 10.1111/j.1530-0277.1998.tb03912.x] [Medline: 9756046]
4. Kelly J, Hoeppner B, Stout R, Pagano M. Determining the relative importance of the mechanisms of behavior change within Alcoholics Anonymous: a multiple mediator analysis. Addiction 2012 Feb;107(2):289-299 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03593.x] [Medline: 21917054]
5. Kelly J. Is Alcoholics Anonymous religious, spiritual, neither? Findings from 25 years of mechanisms of behavior change research. Addiction 2017 Jun;112(6):929-936 [FREE Full text] [doi: 10.1111/add.13590] [Medline: 27718303]
6. Mujcic A, Linke S, Hamilton F, Phillips A, Khadjesari Z. Engagement With Motivational Interviewing and Cognitive Behavioral Therapy Components of a Web-Based Alcohol Intervention, Elicitation of Change Talk and Sustain Talk, and Impact on Drinking Outcomes: Secondary Data Analysis. J Med Internet Res 2020 Sep 01;22(9):e17285 [FREE Full text] [doi: 10.2196/17285] [Medline: 32870162]
7. D'Amico EJ, Houck JM, Hunter SB, Miles JNV, Osilla KC, Ewing BA. Group motivational interviewing for adolescents: change talk and alcohol and marijuana outcomes. J Consult Clin Psychol 2015 Feb;83(1):68-80 [FREE Full text] [doi: 10.1037/a0038155] [Medline: 25365779]
8. Xiao B, Can D, Gibson J, Imel Z, Atkins D, Georgiou P, et al. Behavioral coding of therapist language in addiction counseling using recurrent neural networks. 2016 Presented at: Interspeech; 8-12 Sep 2016; San Francisco p. 908-912. [doi: 10.21437/Interspeech.2016-1560]
9. Challet-Bouju G, Hardouin J, Thiabaud E, Saillard A, Donnio Y, Grall-Bronnec M, et al. Modeling Early Gambling Behavior Using Indicators from Online Lottery Gambling Tracking Data: Longitudinal Analysis. J Med Internet Res 2020 Aug 12;22(8):e17675 [FREE Full text] [doi: 10.2196/17675] [Medline: 32254041]
10. Rodda S, Lubman DI, Dowling NA, Bough A, Jackson AC. Web-based counseling for problem gambling: exploring motivations and recommendations. J Med Internet Res 2013 May 24;15(5):e99 [FREE Full text] [doi: 10.2196/jmir.2474] [Medline: 23709155]
11. Bem D. Self-Perception Theory. Advances in Experimental Social Psychology 1972;6:1-62. [doi: 10.1016/S0065-2601(08)60024-6]
12. Amrhein PC, Miller WR, Yahne CE, Palmer M, Fulcher L. Client commitment language during motivational interviewing predicts drug use outcomes. Journal of Consulting and Clinical Psychology 2003 Oct;71(5):862-878. [doi: 10.1037/0022-006x.71.5.862]
13. Miller W, Benefield R, Tonigan J. Enhancing motivation for change in problem drinking: A controlled comparison of two therapist styles. Journal of Consulting and Clinical Psychology 1993;61(3):455-461. [doi: 10.1037/0022-006X.61.3.455]
14. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. New York: Guilford Press; 2012.
15. Martin T, Moyers T, Houck J, Christopher P, Miller W. Motivational Interviewing Sequential Code for Observing Process Exchanges (MI-SCOPE) Coder's Manual. New Mexico: Center on Alcoholism, Substance Abuse, and Addictions; 2005.
16. Magill M, Apodaca TR, Borsari B, Gaume J, Hoadley A, Gordon REF, et al. A meta-analysis of motivational interviewing process: Technical, relational, and conditional process models of change. J Consult Clin Psychol 2018 Feb;86(2):140-157 [FREE Full text] [doi: 10.1037/ccp0000250] [Medline: 29265832]
17. Miller WR, Rose GS. Toward a theory of motivational interviewing. Am Psychol 2009 Sep;64(6):527-537 [FREE Full text] [doi: 10.1037/a0016830] [Medline: 19739882]

18. Hodgins DC, Ching LE, McEwen J. Strength of commitment language in motivational interviewing and gambling outcomes. Psychol Addict Behav 2009 Mar;23(1):122-130. [doi: 10.1037/a0013010] [Medline: 19290696]
19. Houck J, Moyers T, Miller W, Glynn L, Hallgren K. Motivational interviewing skill code (MISC) 2.5. ELICIT Coding Manual. New Mexico: Center on Alcoholism, Substance Abuse, and Addictions; 2010. URL: https://casaa.unm.edu/download/misc25.pdf [accessed 2021-06-08]
20. Glynn L, Moyers T. Manual for the client language easy rating (clear) coding system: formerly “Motivational Interviewing Skill Code (MISC) 1.1”. Center on Alcoholism, Substance Abuse, and Addictions. 2012. URL: https://casaa.unm.edu/download/CLEAR.pdf [accessed 2021-06-07]
21. Romano M, Peters L. Evaluating the mechanisms of change in motivational interviewing in the treatment of mental health problems: A review and meta-analysis. Clin Psychol Rev 2015 Jun;38:1-12. [doi: 10.1016/j.cpr.2015.02.008] [Medline: 25779459]
22. Swan J. The technical hypothesis of motivational interviewing: an examination of change language in traditional and computer-based MI for disordered gamblers. University of Calgary Library Resources. 2019. URL: https://prism.ucalgary.ca/handle/1880/110167 [accessed 2021-06-07]
23. Magill M, Gaume J, Apodaca TR, Walthers J, Mastroleo NR, Borsari B, et al. The technical hypothesis of motivational interviewing: a meta-analysis of MI's key causal model. J Consult Clin Psychol 2014 Dec;82(6):973-983 [FREE Full text] [doi: 10.1037/a0036833] [Medline: 24841862]
24. Joramo I, Solem S, Romundstad B, Nordahl HM. Change talk and sustain talk in treatment of generalized anxiety disorder: A secondary analysis of cognitive behavioral therapy and metacognitive therapy in adult outpatients. J Behav Ther Exp Psychiatry 2021 Sep;72:101650 [FREE Full text] [doi: 10.1016/j.jbtep.2021.101650] [Medline: 33667828]
25. Richmond RL, Kehoe L, de Almeida Neto AC. Three year continuous abstinence in a smoking cessation study using the nicotine transdermal patch. Heart 1997 Dec;78(6):617-618 [FREE Full text] [doi: 10.1136/hrt.78.6.617] [Medline: 9470884]
26. Silverman K, Svikis D, Wong CJ, Hampton J, Stitzer ML, Bigelow GE. A reinforcement-based Therapeutic Workplace for the treatment of drug abuse: Three-year abstinence outcomes. Experimental and Clinical Psychopharmacology 2002 Aug;10(3):228-240. [doi: 10.1037/1064-1297.10.3.228]
27. Slutske W. Natural Recovery and Treatment-Seeking in Pathological Gambling: Results of Two U.S. National Surveys. American Journal of Psychiatry 2006;163:297-302. [doi: 10.1176/appi.ajp.163.2.297]
28. Oei TPS, Gordon LM. Psychosocial factors related to gambling abstinence and relapse in members of gamblers anonymous. J Gambl Stud 2008 Mar;24(1):91-105. [doi: 10.1007/s10899-007-9071-7] [Medline: 17674163]
29. Okui T. Bulletin board for overcoming gambling addiction-No cracks, no slots, no gambling!. Support Addict Gamblers. 2008. URL: https://9121.teacup.com/fposaka/bbs [accessed 2021-03-10]
30. McGrath DS, Barrett S. The comorbidity of tobacco smoking and gambling: a review of the literature. Drug Alcohol Rev 2009 Nov;28(6):676-681. [doi: 10.1111/j.1465-3362.2009.00097.x] [Medline: 19930023]
31. Grant JE, Steinberg MA, Kim SW, Rounsaville BJ, Potenza MN. Preliminary validity and reliability testing of a structured clinical interview for pathological gambling. Psychiatry Res 2004 Aug 30;128(1):79-88. [doi: 10.1016/j.psychres.2004.05.006] [Medline: 15450917]
32. Yokotani K, Tamura K, Kaneko Y, Kamimura E. Craving for Gambling Predicts Income-Generating Offenses: A Pathways Model of a Japanese Prison Population. J Gambl Stud 2020 Jun;36(2):459-476. [doi: 10.1007/s10899-019-09887-4] [Medline: 31446551]
33. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
34. NTT Communications. A natural language processing and speech processing API platform developed by NTT Communications that utilizes one of the largest Japanese language dictionaries in Japan. COTOHA API. URL: https://api.ce-cotoha.com/contents/index.html [accessed 2020-04-10]
35. Mekala D, Gupta V, Paranjape B, Karnick H. SCDV: Sparse Composite Document Vectors using soft clustering over distributional representations. 2017 May 12 Presented at: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing; September 7-11, 2017; Copenhagen, Denmark p. 659-669. [doi: 10.18653/v1/D17-1069]
36. Rondhuit. URL: https://www.rondhuit.com/download.html [accessed 2020-03-10]
37. Murata M, Isahara H, Nagao M. Resolution of indirect anaphora in Japanese sentences using examples "X no Y (Y of X)". ACL Anthology. URL: https://www.aclweb.org/anthology/W99-0205.pdf [accessed 2021-06-07]
38. Yokotani K, Tamura K. Effects of Personalized Feedback Interventions on Drug-Related Reoffending: a Pilot Study. Prev Sci 2015 Nov;16(8):1169-1176. [doi: 10.1007/s11121-015-0571-x] [Medline: 26051508]
39. Hsu C, Chang C, Lin C. A practical guide to support vector classification. National Taiwan University Repository. URL: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf [accessed 2021-06-07]
40. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/brm.41.4.1149]
41. Kudo T. MeCab: Yet Another Part-of-Speech and Morphological Analyzer. 2006. URL: http://taku910.github.io/mecab/ [accessed 2016-06-24]

42. Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. ICWSM. URL: https://www.semanticscholar.org/paper/Predicting-Depression-via-Social-Media-Choudhury-Gamon/e992de7c53579bd1f23c80be17ca3f6692593ea6 [accessed 2021-06-07]
43. Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Oct 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
44. Baer J, Beadnell B, Garrett S, Hartzler B, Wells E, Peterson P. Adolescent change language within a brief motivational intervention and substance use outcomes. Psychol Addict Behav 2008 Dec;22(4):570-575 [FREE Full text] [doi: 10.1037/a0013022] [Medline: 19071983]
45. Bertholet N, Faouzi M, Gmel G, Gaume J, Daeppen JB. Change talk sequence during brief motivational intervention, towards or away from drinking. Addiction 2010 Dec;105(12):2106-2112. [doi: 10.1111/j.1360-0443.2010.03081.x] [Medline: 20840175]
46. Hodgins D, Currie S, Currie G, Fick G. Randomized trial of brief motivational treatments for pathological gamblers: More is not necessarily better. J Consult Clin Psychol 2009 Oct;77(5):950-960. [doi: 10.1037/a0016318] [Medline: 19803574]
47. Vader A, Walters S, Prabhu G, Houck J, Field C. The language of motivational interviewing and feedback: counselor language, client language, and client drinking outcomes. Psychol Addict Behav 2010 Jun;24(2):190-197 [FREE Full text] [doi: 10.1037/a0018749] [Medline: 20565145]
48. Melville KM, Casey LM, Kavanagh DJ. Psychological treatment dropout among pathological gamblers. Clin Psychol Rev 2007 Dec;27(8):944-958. [doi: 10.1016/j.cpr.2007.02.004] [Medline: 17433853]
49. Pallesen S, Mitsem M, Kvale G, Johnsen B, Molde H. Outcome of psychological treatments of pathological gambling: a review and meta-analysis. Addiction 2005 Oct;100(10):1412-1422. [doi: 10.1111/j.1360-0443.2005.01204.x] [Medline: 16185203]
50. Russ TC, Woelbert E, Davis KAS, Hafferty JD, Ibrahim Z, Inkster B, MQ Data Science group. How data science can advance mental health research. Nat Hum Behav 2019 Jan;3(1):24-32. [doi: 10.1038/s41562-018-0470-9] [Medline: 30932051]
51. Slutske W, Blaszczynski A, Martin N. Sex differences in the rates of recovery, treatment-seeking, and natural recovery in pathological gambling: results from an Australian community-based twin survey. Twin Res Hum Genet 2009 Oct;12(5):425-432. [doi: 10.1375/twin.12.5.425] [Medline: 19803770]

Abbreviations

RBF: radial basis kernel function
SVM: support vector machine

Edited by R Kukafka; submitted 03.09.20; peer-reviewed by K Magnusson, E Marzilli, M Zhao, S Sanchez; comments to author 21.01.21; revised version received 16.04.21; accepted 04.05.21; published 21.06.21.

Please cite as:
Yokotani K
A Change Talk Model for Abstinence Based on Web-Based Anonymous Gambler Chat Meeting Data by Using an Automatic Change Talk Classifier: Development Study
J Med Internet Res 2021;23(6):e24088
URL: https://www.jmir.org/2021/6/e24088
doi:10.2196/24088
PMID:34152282

Original Paper

Interests, Motives, and Psychological Burdens in Times of Crisis and Lockdown: Google Trends Analysis to Inform Policy Makers

Dominik Rotter[1], BSc, MSc; Philipp Doebler[2], Prof Dr; Florian Schmitz[1], Prof Dr

[1]Department of Psychology, University of Duisburg-Essen, Essen, Germany
[2]Statistical Methods in Social Sciences, TU Dortmund University, Dortmund, Germany

Corresponding Author:
Dominik Rotter, BSc, MSc
Department of Psychology
University of Duisburg-Essen
Universitätsstraße 2
Essen, 45141
Germany
Phone: 49 201 183 2173
Email: dominik.rotter@uni-due.de

Abstract

Background: In the face of the COVID-19 pandemic, the German government and the 16 German federal states implemented a variety of nonpharmaceutical interventions (NPIs) to decelerate the spread of the SARS-CoV-2 virus and thus prevent a collapse of the health care system. These measures comprised, among others, social distancing, the temporary closure of shops and schools, and a ban of large public gatherings and meetings with people not living in the same household.

Objective: It is fair to assume that the issued NPIs have heavily affected social life and psychological functioning. We therefore aimed to examine possible effects of this lockdown in conjunction with daily new infections and the state of the national economy on people’s interests, motives, and other psychological states.

Methods: We derived 249 keywords from the Google Trends database, tapping into 27 empirically and rationally selected psychological domains. To overcome issues with reliability and specificity of individual indicator variables, broad factors were derived by means of time series factor analysis. All domains were subjected to a change point analysis and time series regression analysis with infection rates, NPIs, and the state of the economy as predictors. All keywords and analyses were preregistered prior to analysis.

Results: With the pandemic arriving in Germany, significant increases in people’s search interests were observed in virtually all domains. Although most of the changes were short-lasting, each had a distinguishable onset during the lockdown period. Regression analysis of the Google Trends data confirmed pronounced autoregressive effects for the investigated variables, while forecasting by means of the tested predictors (ie, daily new infections, NPIs, and the state of economy) was moderate at best.

Conclusions: Our findings indicate that people’s interests, motives, and psychological states are heavily affected in times of crisis and lockdown. Specifically, disease- and virus-related domains (eg, pandemic disease, symptoms) peaked early, whereas personal health strategies (eg, masks, homeschooling) peaked later during the lockdown. Domains addressing social life and psychosocial functioning showed long-term increases in public interest. Renovation was the only domain to show a decrease in search interest with the onset of the lockdown. As changes in search behavior are consistent over multiple domains, a Google Trends analysis may provide information for policy makers on how to adapt and develop intervention, information, and prevention strategies, especially when NPIs are in effect.

(J Med Internet Res 2021;23(6):e26385) doi:10.2196/26385

KEYWORDS

coronavirus; Google Trends; infodemiology; infoveillance; pandemic; information search; trend; COVID-19; burden; mental health; policy; online health information

Introduction

On March 11, 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a pandemic, as almost all countries of the world were affected [1]. One month earlier, the WHO announced the official names of this coronavirus and the disease it causes as SARS-CoV-2 and COVID-19, respectively [2]. As of October 3, 2020, there have been 296,958 reported cases in Germany [3] and 34,495,176 reported cases worldwide [4].

With increasing case numbers and death counts comes the necessity to monitor and slow the further spread of the pandemic [5]. There is strong evidence that human-to-human transmission is the reason the initial outbreak in Wuhan, China, became global; thus, it is imperative to prevent secondary infections [6] (eg, by means of restricting air traffic [7] and other nonpharmaceutical interventions [NPIs]) in the absence of a vaccine [8]. NPIs are used to pursue two strategies—mitigation and suppression of further spread—which can be achieved by a combination of multiple population-wide measures, such as social distancing by the entire population; home isolation and quarantine of cases and their household members; banning mass gatherings, and closing schools and universities [8,9]. The implementation of these measures seems to slow the spread of the pandemic, thus preventing thousands of deaths [9,10]. As data for monitoring disease spread is difficult to obtain in the early stages of a pandemic, the use of internet data has become increasingly valuable, with the emergence of two research disciplines: infodemiology and infoveillance [11]. Although the former is a research discipline and methodology studying the determinants and distribution of health information [12], the latter is defined as the longitudinal tracking of infodemiology metrics with the goal of surveillance and trend analysis [13]. Both can help policy makers in developing prevention and information strategies. As information influences knowledge, behavior changes, and health outcomes [11,14], the internet can be used as a handy tool to monitor public search behavior [15-17] or public awareness [18] of health care issues, allowing policy makers to quantify knowledge translation gaps [12], identify misleading information circulation [19,20] or possible emerging local outbreaks [21], assess people’s needs [22], derive suitable interventions [23-25], and develop enhanced communication strategies especially for at-risk subpopulations [26].

To inform policy makers about crisis-related developments, researchers have examined changes in search behavior during the COVID-19 pandemic [27,28]. Google Trends data was used to monitor public interest in the pandemic from January to March 2020 by conceptualizing the demand for information as a single search term (“coronavirus”), showing worldwide peaks in people’s interest at the end of January as well as shortly after COVID-19 was declared a pandemic. Pearson correlations indicated a moderate to high relationship between different countries regarding search behavior [29], although some studies (eg, Hu et al [18]) suggest differences during the first peak between countries and even various subregions of countries. Typically, announcements of the first local COVID-19 case led to a widespread increase in search interest as people sought information about the COVID-19 pandemic [30]. There is evidence, though, that COVID-19 terms trending in Italy elicited differential peaks, showing that regions most heavily affected in terms of infections and deaths were not always the first to display changes in search interest [19], as local and international events are capable of influencing public interest [31]. However, the number of recent COVID-19 infections and deaths is associated with changes in public interest, which becomes apparent by maximum correlations when a lead of approximately –11.5 days is applied [32]. This lagged nature of relationships has also been found for other COVID-19–related search tags (eg, the frequency of search queries for loss of smell [33,34] and taste [34,35] as key symptoms of COVID-19, or public interest in “pneumonia” [36] or “high temperature”, “cough”, and “diabetes” [37], as well as “insomnia”, a key signal for mental distress [38]). The positive relationship seems to be consistent across different social media platforms and search engines, such as Baidu and Sina Weibo Indices [36], as well as with other diseases (eg, the flu outbreak in 2004 [12]). However, Mavragani [39], who used Google Trends data with the search tag “coronavirus” to study European countries (ie, Italy, Spain, France, Germany, and the United Kingdom) found moderate to high positive Pearson correlations between search queries and the number of total and daily new cases for all countries except Italy. An in-depth analysis of multiple time frames indicated a declining trend of the Pearson correlation coefficient over time, highlighting an early interest phenomenon. Overall, the increase of attention is typically short-lived [18,29], even when accompanied by NPIs issued by policy makers [30].

Given that Google Trends data can indicate population health literacy, Google Trends queries for “wash hands” and “face mask” were found to correlate with a lower spreading rate of COVID-19 cases in 21 countries: the logarithmic increase of case numbers correlated negatively with the number of days with high search volume [40]. Google Trends data could moderately predict new cases, although predictions were not found to be very precise [41]. Nonetheless, COVID-19 symptoms showed some predictive value, especially when virus diffusion was monitored by predicting the number of deaths [42]. Even real-time forecasting of the COVID-19 outbreak has been shown to be possible [43,44].

However, changes in search behavior are not only limited to disease symptoms (eg, “anosmia” [45]). Previous research has similarly shown that COVID-19 has affected the social, economic, and psychological well-being of humanity, pervading broad domains of social and daily life, such as COVID-19 awareness [29,32], psychological stress (eg, “anxiety”, “depression”, “therapy” [46]), food supply (eg, “sourdough”, “restaurants”, “baking” [46]), economic stressors (eg, “mortgage”, “homeloan”), social stressors (eg, “divorce”, “liquor”), and treatment seeking (eg, “cognitive therapy”, “counselling” [25]), as well as quarantine and even conspiracy theories [30]. An analysis of Twitter data identified four main domains of interest for people: the origin of the virus, its sources, its impact on people, countries, and economies, as well as ways of mitigating the risk of infection. The latter suggests that Twitter users are keen to learn and also share their knowledge with friends and followers, while also discussing the negative consequences of the COVID-19 pandemic and its emotional and psychological impact [22]. Further research found changes in stress-related and food-related searches, indicating that the outbreak may have had a significant impact on both people's stress levels and daily routines [46].

Despite a growing body of research studying search behavior regarding COVID-19 and its symptoms in conjunction with infection and death counts, there seem to be limited research activities targeting the effects of the issued NPIs. One early study used Google Trends data from January to March 2020 to examine the impact of the pandemic and the issued NPIs on people's concerns, indexed by changes in search frequency about mental health issues as well as financial and work-related consequences. They found an increase in financial and work-related queries, indicating that people are well aware of further consequences, such as job loss. There is also some evidence that queries for "depression" and "suicide" decreased immediately after the announced pandemic, but increased shortly after the first wave [25]. Another study stressed the potential damage to a population's well-being as a consequence of the lockdown. It used Google Trends data to test for changes in people's well-being and found a substantial increase in search intensity for boredom in European countries and the United States, which was accompanied by a higher search volume for loneliness, worry, and sadness, while searches for stress, suicide, and divorce declined [47].

Since previous research has shown that the COVID-19 pandemic has resulted in changes in public interest [18,29,30] beyond the expected search behavior regarding COVID-19–related symptoms [22], we conducted this study to investigate the effects of the pandemic and the corresponding lockdown using a broad set of psychological domains and interests. As Google Trends data can function as a proxy for psychological states and behaviors [46], search interests may indicate changes in social and daily life. Our 27 domains of interest range from COVID-19 restrictions to online shopping, sexual interests, and psychosocial impact. The German government and the 16 German federal states issued a cluster of NPIs to decelerate the spread of COVID-19. These measures comprised, among others, social distancing, the temporary closure of shops and schools, and banning large public gatherings and meetings with people not living in the same household [5]. As prior studies suggested, it seems fair to assume that all measures heavily affected social life and psychological functioning [25,46-48]. Thus, we expected significant changes in search behavior during the emerging pandemic and the subsequently issued lockdown period for all observed domains. Further, daily new infections were tested as a predictor, given that infection and death counts have been shown to be associated with an increase in public interest [32,35]. Additionally, the state of the national economy was investigated as a predictor, as it may affect people's employment situation and related psychological anticipations, worries, and the need to search for COVID-19–related information. In turn, changes in search queries and their relevant predictors can be used to inform policy makers, allowing them to tailor information and health prevention strategies [23-25].

Methods

Psychological Domains and Indicators of Interest

As preregistered [49], we planned to examine 27 psychologically meaningful domains that may possibly have been affected by the COVID-19 pandemic, the NPIs, or associated economic developments. For each domain, at least three search terms were selected as indicators. For instance, the domain "pandemic disease" comprised "coronavirus", "covid sars", or "covid-19" as Google search tags. Descriptions of all domains are given in Table 1. For all search terms, refer to Multimedia Appendix 1.

Table 1. All 27 domains and their relevant description.

Domain	Description
Pandemic disease	People's interest in the virus and the disease it causes
Health care institutions	German and international institutions as well as experts providing information on the pandemic
Political leaders	Names of leading German politicians whose responsibilities comprise dealing with the pandemic
Infection	Information concerning the rapidity and method of virus spread
Symptoms	Possible symptoms experienced by patients with COVID-19 indicating a virus infection
Information seeking	Sources and methods people could use to obtain an overview of the development of the pandemic
Testing	Questions pertaining to the capacities of testing, including locations and general opportunities to have oneself tested
Hand hygiene	Possibilities of keeping one's hands virus-free
Mask	Interest in the different types of masks, their effectiveness, and the current regulations on wearing a mask
Disinfectant	Possible substances used for disinfection and recommendations for appropriate disinfection
Convenience goods	Groceries or household goods that were nearly sold out due to panic buying and hoarding
Vaccination	Interest in vaccination possibilities and fear of a compulsory vaccination
Parenting and childcare	Challenges parents have to deal with during the pandemic
COVID-19 restrictions	Information about the current restrictions and rules of conduct
COVID-19 relaxations	Information about current and future possibilities and relaxations of the issued restrictions
Economic impact	Economic shortcomings and development as consequences of the COVID-19 restrictions
Sexual interest	Changes in sexual interest due to the COVID-19 crisis
Social life	Possibilities for people to socialize while following COVID-19 restrictions such as social distancing
Homeschooling	Information about new ways of educating students at home while schools were closed
Business communication	Different ways for employees to stay in contact with their colleagues and to be able to work from home
Hobbies and sports	Activities that became popular during the lockdown as they could be done without breaking restrictions
Renovation	Renovation activities that could be done at home during the pandemic
Online shopping	Names of common online sellers
Dispatching	Terms and information about the shipment of items
Psychosocial impact	Negative consequences that arose due to the COVID-19 restrictions
Conspiracy theories	Several conspiracy theories related to the source and spread of the pandemic
Government support	Terms and information about financial aid due to closures and severe restrictions

Retrieving Google Trends Data

Each domain was entered in the Google Trends database to reveal the number of Google search queries for a given time period and geolocation using the *gtrends* package [50] for the statistical language *R* [51]. Before downloading, we thoroughly studied the framework for using Google Trends data, which highlights the choice of region, time period, and geolocation [11]. As Google allows one to set different geolocations at the country and federal state level within a specified period, we retrieved Google Trends data separately for each search tag for 17 geolocations (ie, Germany and its 16 federal states) for the period from January 6 to September 1, 2020. We chose this period to allow for a sufficient period before the first cases occurred in Germany as a baseline estimate. By default, Google standardizes the queries per day by the maximum of queries within the given period, multiplying all values by 100. High interest in a query is expressed by 100, whereas minor interest or missing data on queries is expressed by 0. Values below 1 were recoded as 1.

We retrieved Google Trends data for all domains for all 17 geolocations except for 32 search tags falling into 16 domains. An overview of the domains and search tags for which we could retrieve Google Trends data is provided in Multimedia Appendix 1. For the analysis, we used only keywords available for all federal states. As there were differences between the retrieved time series for the geolocation "DE" for Germany and the aggregated time series across all federal states, we decided to use the aggregated time series for further analysis.

Retrieving Daily New Infections Data

As the Robert Koch Institute (RKI), a German federal government agency and research institute responsible for disease control and prevention, issued daily situation reports concerning COVID-19 [52] even before the official declaration of the pandemic, we made use of the online dashboard application programming interface (API) to retrieve daily data, comprising infection and death counts for each federal state as well as Germany as a whole from January 6 to September 1, 2020. Maps of the German federal states together with 7-day incidence rates are provided in Multimedia Appendix 2, along with summary statistics for, for example, population density, cumulative infections, and incidence rates in Multimedia Appendix 3.

Retrieving Economic Data

The truck toll mileage index (TTMI) provides an early and sensitive day-to-day proxy for the state of industrial production. It is computed by using process data of the toll system for trucks in Germany [53]. According to the German Federal Statistical Office (Statistisches Bundesamt), the TTMI can serve as a highly accurate indicator of the economic state with a lag of 5 to 9 days. We retrieved the data for the period from January 6 to September 1, 2020.

Retrieving NPI Data

In the face of the COVID-19 pandemic, the German government and its federal states issued 14 NPIs to curb the spread of COVID-19. The measures included limitations regarding leaving home without reason (ie, only for grocery shopping or a doctor’s appointment), recommendations to keep social distance (>1.5m), the obligation to wear a mask in stores and when using public transport, closure of nonessential shops, closure of hairdressers and close contact services (eg, massage studios), closure of nonfood shops (eg, bicycle and hardware stores), closure of zoos, the prohibition of public demonstrations; closure of schools, playgrounds, kindergartens, and day care facilities; closure of religious institutions, and prohibition of meeting in public with persons not belonging to the same household. The Leibniz Institute for Psychology (ZPID) provided coded daily data for all relevant NPIs with their onset, separated by federal state, from March 8 to June 26, 2020 [5]. We will subsequently refer to this period as the lockdown period.

Data Reduction and Preprocessing

Within each domain, a time series factor analysis (TSFA) [54] was used for Germany to test if one hypothesized dominant factor accounts for the covariation of indicators in the respective domain. First, we extracted the moving average for 7 days (the so-called trend) for all search terms for a domain. Second, we z-standardized the time series to attain loadings, which can be interpreted as correlations ($-1 \leq \lambda \leq 1$), and estimated the factor analysis with a single assumed factor. We used only the trend information of time series as it represents the long-term process and behavior of the time series [55]. Indicator variables were selected when they had positive loadings on the first factor, thereby iteratively excluding indicators with negative loadings. We excluded 20 search terms falling into 10 domains (Multimedia Appendix 1).

For all domains, the scree plots resulting from the selection process are presented in Multimedia Appendix 4. Next, factor scores were calculated using the loadings on the first factor as weights to aggregate a single time series for each keyword per federal state, retrieving a mean time series for each domain. We then deseasonalized the time series for each domain, supposing an additive model with weekly seasonality for every time series.

Since the NPIs were issued at different times, we used TSFA to identify factors that describe their covariation. We differenced every time series, so they corresponded with daily change scores, prior to computing the TSFA. Scree plots could be interpreted as indicating a two- or three-factorial structure. The two-factor model could be more consistently interpreted, yielding one factor of “regulation of outdoor activities” (NPI Outdoor) and another factor of “regulation of social life” (NPI Social). Corresponding with the selection criteria applied to Google Trends data, we excluded indicator time series with negative loadings. We calculated factor scores using the loadings as weights to aggregate single time series for each NPI, resulting in a single mean time series for each NPI factor. Details concerning indicators and their respective loadings are given in Table 2.

Table 2. TSFA loadings for the two extracted NPI factors, regulation of outdoor activities (NPI Outdoor) and regulation of social life (NPI Social), for aggregated time series data[a].

NPI	Regulation of outdoor activities (NPI Outdoor)	Regulation of social life (NPI Social)
Prohibition to leave home	N/A[b]	.91
Social distancing	.31	.74
Rule to wear a mask in public	.06	N/A
Closure of nonessential shops	.28	N/A
Closure of hairdressers and close contact services	.02	.58
Closure of nonfood shops	.39	.22
Closure of zoos	.99	N/A
Prohibition of public demonstrations	.32	N/A
Closure of schools	.41	.05
Closure of religious institutions	.67	.05
Prohibition to meet with nonhousehold members	N/A	.73
Prohibition to meet with more friends	.14	.69
Closure of shops	.93	N/A
Closure of kindergartens and day care	.48	.10

[a] Indicators with negative loadings were omitted from the factor scores. NPI: nonpharmaceutical intervention. TSFA: time series factor analysis.

[b] N/A: not applicable.

Visual Inspection and Change Point Analysis

Each aggregated time series per domain and federal state was carefully screened in conjunction with the time series for daily new infections, as well as the economic state and the two NPI factors “regulation of outdoor activities” and “regulation of social life.” As we assumed structural breaks, indicating a major change in the underlying processes of the time series, we proceeded with a change point analysis with an intercept-only model to identify changes in the level for the lockdown period [56].

Time Series Regression Analysis

For regression analysis, we conceptualized search behavior over time for each federal state as a criterion variable in linear mixed models, testing intercept-only models to identify a possible hierarchical structure as it would be indicated by a substantial intraclass correlation (ICC). Since there were no meaningful ICCs (all ICC<0.05, as preregistered), we proceeded with models for Germany only. We used daily new infections, the NPI factors NPI Outdoor and NPI Social, as well as the TTMI as predictors. Criterion and predictor variables were z-standardized before linear modelling. To account for the autocorrelation in time series, we used an autoregressive integrated moving average (ARIMA) approach to fit the models [55,57]. ARIMA models are a form of generalized least squares models with three parameters, namely *p*, *d*, and *q*. The *p* parameter denotes the order of the autoregressive part AR(*p*). In an autoregressive model, the values in a time series Y_t are predicted by its lagged values Y_{t-p}. For a lag of 1, the AR(1)-model is equal to equation 1.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t \quad \mathbf{(1)}$$

For the moving average part, the parameter *q* denotes the order of previous error values ε_{t-q} used to predict the time series Y_t. An MA(*1*)-model is equal to equation 2.

$$Y_t = \beta_0 + \beta_1 \varepsilon_{t-1} + \varepsilon_t \quad \mathbf{(2)}$$

The *d* parameter stands for “integrated” and denotes the order of differencing that was applied to the time series before an autoregressive moving-average (ARMA) analysis. Differencing is used to adjust for seasonality in the time series and can be repeated multiple times. In summary, an ARIMA model with the parameters p=1, d=1, and q=1 with a predictor X_1 and $D_t = Y_t - Y_{t-1}$ is equal to equation 3.

$$D_t = \beta_0 + \beta_1 X_1 + \beta_2 \varepsilon_{t-1} + \beta_3 D_{t-1} + \varepsilon_t \quad \mathbf{(3)}$$

In time series regression, predictors are frequently allowed to exert lagged effects. To estimate this lag period, the time parameter is adjusted (shifting the time series) to maximize predictive power. As we were only interested in predicting Google Trends data, we focused on shifting predictors forward from 0 to 21 days. In this case of shifting a time series forward, the value of a time series (Y_t) is replaced by its successive value (Y_{t+1}), which is called a lead. We refrained from using the Bayesian information criterion [58] for model selection as preregistered [49], since iteratively shifting time series resulted in different combinations of ARIMA parameters, rendering a meaningful interpretation of the respective fit values impossible. Therefore, we selected the lag period for which the predictor exerted a maximal correlation with the criterion variable. The cross-correlation functions for all relevant predictors (ie, daily new infections, NPI factors NPI Outdoor and NPI Social, TTMI) are provided in Multimedia Appendix 5.

All statistical analyses were performed using the statistical language *R* [51] with the package *strucchange* [56] for identifying structural breaks and the package *lme4* [59] for linear modeling.

Results

Cross-Correlation Between Domains

Search interests for all 27 domains were moderately to highly correlated (Figure 1, lower triangle). Exceptions can be noticed for the domains *parenting and childcare*, *COVID-19 relaxations*, *sexual interest*, *renovation*, and *government support*. To identify possible lagged relations between domains, we shifted one time series in the range from 0 to 21 days. The maximum correlations obtained within this time-lagged range were highly comparable to the concurrent correlation. Hence, the overall pattern of correlations across domains was hardly affected (Figure 1, upper triangle).

Figure 1. Heat map of cross-correlations for all 27 domains for Germany only. The lower triangle shows concurrent cross-correlations. The upper triangle shows maximal correlations obtained when the time series of one domain was shifted across a range from 0 to 21 days relative to the time series of the corresponding other domain. CCF: cross-correlation function.

Visual and Change Point Analysis

All domains showed pronounced changes in people's search behavior in the corresponding period from January 6 to September 1, 2020, although the changes were mostly short-lived (Figures 2 and 3). Many domains showed different onsets during the COVID-19 crisis. Disease- and virus-related domains (ie, *pandemic disease*, *health care institutions*, *infection*, *symptoms*, *testing*, *hand hygiene*, *disinfectant*, *parenting and childcare*, *COVID-19 restrictions*, and *economic impact)* show an early onset, but searches decreased quickly. Public interest in political leaders peaked multiple times, while the domains *vaccination*, *social life*, *homeschooling*, *hobbies and sports*, *psychosocial impact*, and *dispatching* seem to be of higher, long-term interest. Interest in other domains seemed to be related to the onset of the NPIs (eg, the domain *mask* peaked when the obligation to wear masks in shops and public transport was announced). Interest in *government support* (ie, financial support programs provided by the national government) increased in later stages of the lockdown, while *COVID-19 relaxations* and *conspiracy theories* attracted attention primarily in the middle of the lockdown phase. For each domain, noticeable changes in the level (ie, mean for a chosen period of time) indicate significant changes in public interest during the examined lockdown period from March 8 to June 26, 2020, in Germany.

Figure 2. Z-standardized Google Trends for the first 14 domains. The red function indicates the trend for Germany, while the grey functions correspond with the trends for the 16 federal states. The grey shaded area represents the lockdown period in Germany from March 8 to June 26, 2020. Change points, indicating level changes, are represented by the vertical blue dashed lines. The intensity of daily new infections (Cases), countermeasures (NPI Social and NPI Outdoor), and the proxy for the state of economy (TTMI) is plotted below. NPI: nonpharmaceutical intervention; TTMI: truck toll mileage index.

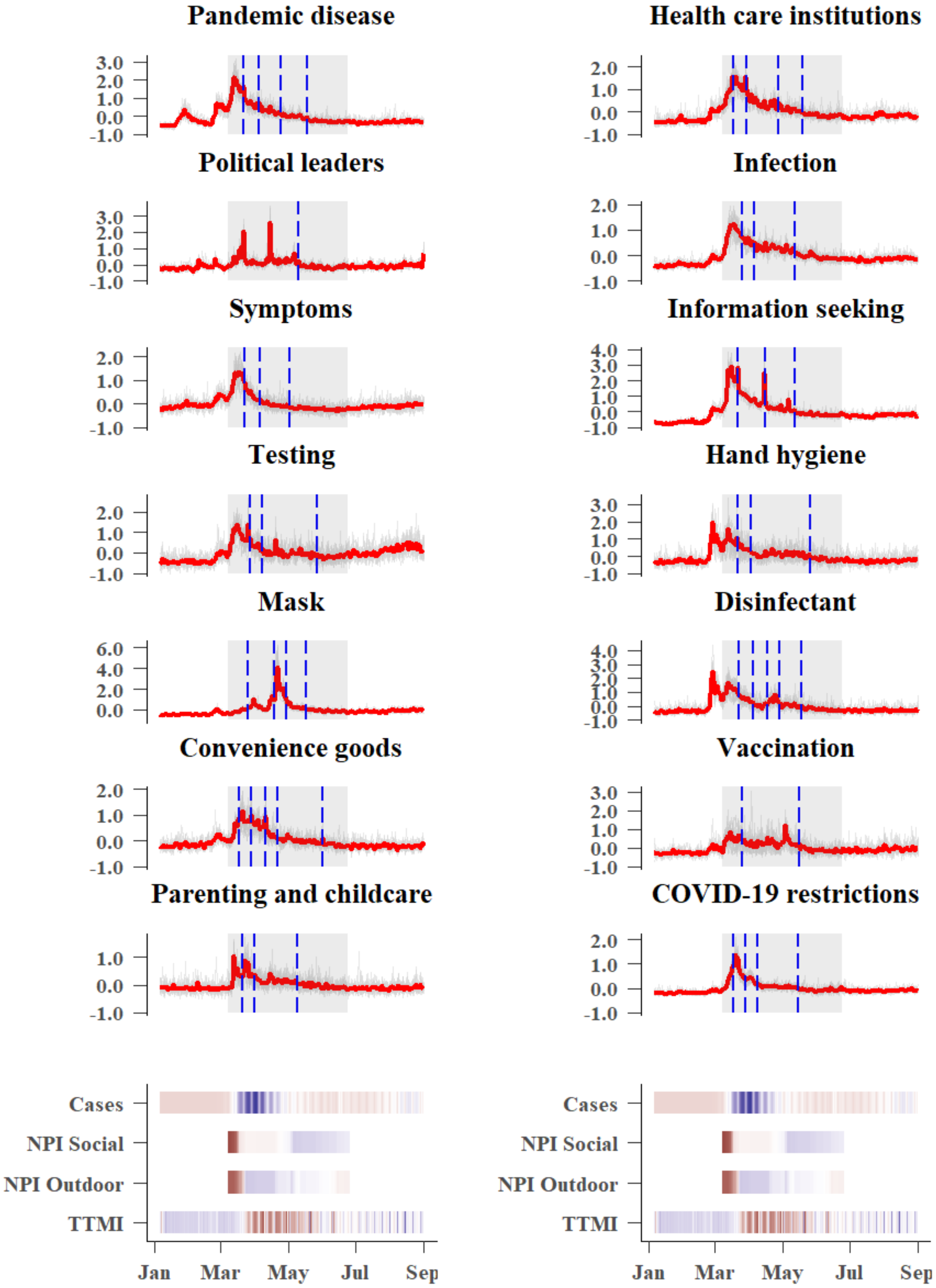
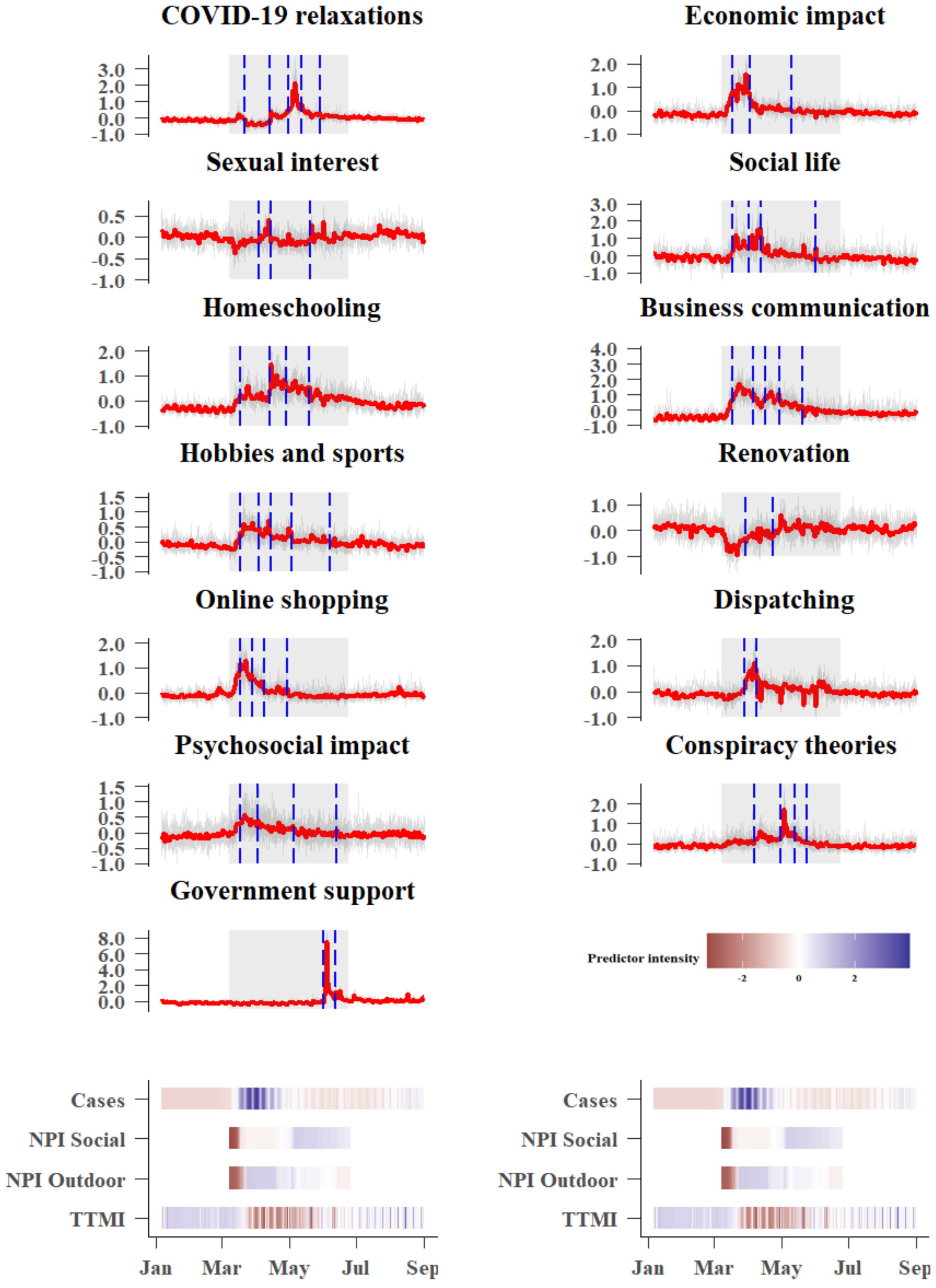


Figure 3. Z-standardized Google Trends for 13 additional domains. The red function indicates the trend for Germany, while the grey functions correspond to the trends for the 16 federal states. The grey shaded area represents the lockdown period in Germany from March 8 to June 26, 2020. Change points, indicating level changes, are represented by the vertical blue dashed lines. The intensity of daily new infections (Cases), countermeasures (NPI Social and NPI Outdoor), and the proxy for the state of economy (TTMI) are plotted below. NPI: nonpharmaceutical intervention; TTMI: truck toll mileage index.



Time Series Regression Analysis

We conceptualized search interests in all 27 domains as criterion variables in ARIMA regression analyses and entered four predictors—namely, daily new infection cases, both NPI factors (NPI Outdoor and NPI Social), and the TTMI as a proxy for the economic state. All analyses were conducted for Germany as a whole, as low ICCs indexed no meaningful heterogeneity between federal states. Table 3 summarizes the results for the regression part. Overall, all predictors showed a different pattern and, thus, predictive value for different domains. Although new infections seemed to have a delayed effect on public interest (see the “Lead” column), the NPI factors and TTMI had their highest effect directly at lag 0. Almost all models benefitted from the ARIMA approach, as shown in Table 4. Most domains needed differencing to account for nonstationarity. Additionally, to handle the correlated residual structure, almost all domains were estimated with an autoregressive or moving average component.

Table 3. ARIMA regression results for the regression part with standardized betas for the intercept and all relevant predictors (infection cases, both NPI factors, and the TTMI), separate for all domains[a].

Domain	Regression term											
	New infections			NPI Outdoor			NPI Social			TTMI		
	Lead	β_1	*P* value	Lead	β_2	*P* value	Lead	β_3	*P* value	Lead	β_4	*P* value
Pandemic disease	–12	.15	.03	0	–.22	.21	0	–.26	.09	0	–.02	.27
Health care institutions	–9	.23	.02	0	.25	.07	–13	.60	<.001	–13	–.09	.03
Political leaders	–5	.26	.16	0	.01	.99	–11	.26	.40	–11	–.18	.06
Infection	–7	.09	.24	0	.14	.40	–12	.61	<.001	–12	–.05	.09
Symptoms	–11	.10	.03	0	.03	.82	0	–.47	<.001	0	–.03	.17
Information seeking	–11	.07	.58	0	.38	.24	0	–.96	<.001	0	–.08	.04
Testing	–13	.08	.58	0	–.09	.76	0	–.27	.33	0	–.11	.03
Hand hygiene	–13	.24	.04	0	–.17	.37	0	–.21	.22	0	–.03	.52
Mask	–18	–.03	.86	–21	.77	.21	0	–.04	.87	0	–.06	.03
Disinfectant	–12	.03	.77	0	–.04	.85	0	–.21	.32	0	.01	.90
Convenience goods	–4	.18	.04	0	.29	.01	–9	.42	<.001	–9	–.06	.06
Vaccination	–5	.05	.72	0	–.33	.16	–13	1.18	<.001	–13	–.05	.44
Parenting and childcare	0	–.32	.02	0	–.25	.18	0	–.17	.42	0	.07	.38
COVID-19 restrictions	–6	.07	.27	0	.44	.02	–10	.57	<.001	–10	–.04	.19
COVID-19 relaxations	–1	–.13	.35	–2	.65	<.001	–15	.18	.70	–15	.02	.56
Economic impact	–1	.48	<.001	–10	–.05	.94	–8	.36	.07	–8	–.02	.70
Sexual interest	–12	–.12	.58	0	.24	.47	0	.05	.86	0	.16	.10
Social life	–4	.21	.15	–10	–.49	.18	–3	.57	<.001	–3	–.12	.07
Homeschooling	–18	–.07	.75	0	.49	.27	0	.26	.48	0	–.17	<.001
Business communication	0	.34	<.001	–10	–.09	.88	–4	.34	.06	–4	–.09	.06
Hobbies and sports	0	–.16	.14	–10	–.04	.93	–3	1.07	<.001	–3	–.05	.27
Renovation	–11	–.68	<.001	0	.01	.99	0	.06	.68	0	–.01	.91
Online shopping	–3	.30	<.001	0	.47	<.001	–9	.49	<.001	–9	–.10	.05
Dispatching	0	.09	.53	0	–.03	.89	0	.27	.19	0	–.43	<.001
Psychosocial impact	–3	.26	.09	0	.29	.11	–9	.40	.02	–9	–.05	.48
Conspiracy theories	–21	–.03	.88	–21	.25	.82	–1	.31	.33	–1	–.07	.40
Government support	–7	.11	.59	0	.19	.28	–13	–.41	.27	–13	.23	.01

[a] Lead corresponds to the forward-shifted time series for the corresponding number of days. Zero values indicate the original time series without shift. ARIMA: autoregressive integrated moving average; NPI: nonpharmaceutical intervention; TTMI: truck toll mileage index.

Table 4. ARIMA regression results for the ARIMA model part with standardized betas for the autoregressive and moving average components[a].

Domain	ARIMA (*p*,*d*,*q*)			ARIMA term							
				Autoregressive component				Moving average component			
	p	*d*	*q*	AR(1)	*P* value	AR(2)	*P* value	MA(1)	*P* value	MA(2)	*P* value
Pandemic disease	1	1	1	–.94	<.001	N/A[b]		.85	<.001	N/A	
Health care institutions	0	1	1	N/A		N/A		–.64	<.001	N/A	
Political leaders	2	0	0	.32	<.001	.21	.04	N/A		N/A	
Infection	0	1	1	N/A		N/A		–.28	.02	N/A	
Symptoms	1	1	3	.61	<.001	N/A		–1.03	<.001	.40	.04
Information seeking	0	1	0	N/A		N/A		N/A		N/A	
Testing	0	1	1	N/A		N/A		–.26	.03	N/A	
Hand hygiene	0	1	1	N/A		N/A		–.57	<.001	N/A	
Mask	1	1	1	–.59	<.001	N/A		.85	<.001	N/A	
Disinfectant	0	1	0	N/A		N/A		N/A		N/A	
Convenience goods	0	1	2	N/A		N/A		–.53	<.001	–.22	.06
Vaccination	0	1	1	N/A		N/A		–.52	<.001	N/A	
Parenting and childcare	0	0	1	N/A		N/A		.40	<.001	N/A	
COVID-19 restrictions	1	0	2	.97	<.001	N/A		–.41	<.001	.30	<.001
COVID-19 relaxations	1	0	0	.87	<.001	N/A		N/A		N/A	
Economic impact	0	1	1	N/A		N/A		–.50	<.001	N/A	
Sexual interest	0	1	1	N/A		N/A		–.70	<.001	N/A	
Social life	0	1	2	N/A		N/A		–.57	<.001	–.34	<.001
Homeschooling	0	1	0	N/A		N/A		N/A		N/A	
Business communication	0	1	1	N/A		N/A		–.48	<.001	N/A	
Hobbies and sports	1	1	1	.28	.03	N/A		–.85	<.001	N/A	
Renovation	1	0	0	.41	<.001	N/A		N/A		N/A	
Online shopping	0	1	1	N/A		N/A		–.68	<.001	N/A	
Dispatching	1	0	0	.56	<.001	N/A		N/A		N/A	
Psychosocial impact	0	1	1	N/A		N/A		–.89	<.001	N/A	
Conspiracy theories	0	1	1	N/A		N/A		–.46	<.001	N/A	

[a] The ARIMA model parameters *p*, *d*, and *q* represent the order of the autoregressive components, the order of differencing of the original time series, and the order of the moving average components, respectively. ARIMA: autoregressive integrated moving average.

[b] N/A: not applicable.

Discussion

Principal Results

This study was conducted to test possible effects of the COVID-19 pandemic, in conjunction with lockdown NPIs, daily new infections, and the state of the national economy, on people's social and psychological interests, going beyond previous research examining only a few domains for a limited timespan [33,35-37]. Specifically, we tested search interests for 27 broad domains capturing social life and psychological functioning. To this end, 249 search terms from the Google Trends database were used as observed indicators. Factor scores obtained in time series factor analyses were used to overcome the challenges of unreliability and specificity of single indicator variables. Search interests in all investigated domains were found to be heavily affected by the COVID-19 pandemic. Confirming previous findings, increases in search interests were mostly short-lived [18,29], indicating possible knowledge translation gaps, which could be overcome by adapting public information strategies [23-25].

While disease- and virus-associated domains (ie, *pandemic disease*, *health care institutions*, *infection*, *symptoms*, *hand hygiene, disinfectant*, and *COVID-19 restrictions*) showed early increases in public interest, the domains *homeschooling* and *mask* were not triggered by first reports but peaked later, when the lockdown measures had been issued for some time already. Searches for the names of *political leaders* peaked multiple times, whereas the domains *vaccination*, *social life*, *hobbies and sports*, *psychosocial impact*, and *dispatching* revealed more long-term interest. This is in line with previous research, which has shown that people responded to the first reports of infections by seeking out information about COVID-19, as measured by searches for coronavirus, coronavirus symptoms, and hand sanitizer [30]. Contrary, personal health strategies (ie, mask, medication) and community-level policies (ie, school closures, testing) were found to peak at later stages of the pandemic [30]. Additionally, differences between domains and their delayed occurrence have been reported multiple times [19,30,46].

We did not find meaningful differences between German federal states for any of the studied domains. This corresponds with high correlations for the single search term "coronavirus" for multiple countries in previous research [29]. However, other research has shown differences in the onset of COVID-19 search terms in Google Trends, even for subregions within countries [18]. Hence, we assume that the highly comparable search interests observed across German federal states could reflect, in part, that German federal states issued comparable clusters of NPIs simultaneously [5].

Generally, cross-correlations between domains were found to be highest for nonshifted time series, which we tentatively interpret as indicating that the announcement of NPIs immediately influences public interest. This potentially reveals a decisive point worth considering when developing information strategies.

Change point analyses indicate level changes in public interest over the lockdown period for all domains. However, there was no conclusive evidence that change points for domains consistently coincide with the onset of the NPIs. This suggests that changes observed in search interests for domains are not directly related with the onset of NPIs. Nevertheless, certain domains characterized by a high cross-correlation (eg, *pandemic disease* and *health care institutions*) revealed a comparable pattern of change points (see Figure 1 for all cross-correlations and Figures 2 and 3 for the time series with change points). Change points may thus be prone to changes within and between time series but are not sensitive enough to coincide with the onset of new infection cases or the onset of NPIs. They could provide circumstantial evidence of co-occurrence of events, but they lack sensitivity to identify temporarily lagged effects.

To shed light on possibly lagged effects, all domains were subject to time series regression. In line with previous evidence [32,36,37] that search interests for COVID-19–related search tags follow increases in infection rates after 8-14 days, daily new infections were a leading predictor in the current study as well. This confirmed a comparably lagged pattern, especially for the domains *pandemic disease*, *health care institutions*, *symptoms*, and *hand hygiene* within the aforementioned lead (see Table 3). A rise in new infection cases also predicted search interest for *convenience goods*, *economic impact*, and *business communication* but was related to reduced interest in *parenting and childcare* and *renovation*. As the lockdown has been shown to considerably affect social life and people's routines [25,46,47], we chose the two NPI factors capturing regulations pertaining to outdoor activities and social life as additional predictors. Both factors appeared to be differentially related with domains. The factor capturing regulation of outdoor activities was positively related with search interests for *convenience goods*, *COVID-19 relaxations*, and *online shopping*. The factor capturing regulation of social activities predicted search interests with the same direction for the same domains as the factor capturing regulation of outdoor activities, and additionally *health care institutions*, *infection*, *vaccination*, *social life*, *hobbies and sports*, and *psychosocial impact*. Further, it was negatively related with the domains *symptoms* and *information seeking*. Overall, our results indicate that NPIs targeting the regulation of social life may have more impact on people's behavior and routines.

The proxy variable for the state of economy (TTMI) revealed relatively consistent negative relations with the investigated search interests of arguably low magnitude. The direction of this effect is plausible, as it indicates increasing interest as the economy declines. However, the small magnitude of these effects suggests that the economic situation had a rather small impact on people's social and psychological interests in the first lockdown period.

Limitations

This study used the information available in public databases and Google Trends as a proxy for people's psychological interests, motives, and concerns. Given that inferences were drawn only from individuals with access to the internet who use the Google search engine, results could be biased to some extent (eg, because older adults or low-income people could be underrepresented in the sample). However, we deem this effect to be rather small, as 95% of households in Germany have access to the internet [60]. Further, there is evidence showing that information drawn from different search engines and social media platforms is highly comparable [36], hence, results can be expected not to be specific for the Google search engine only.

We investigated whether changes in search interests were affected by the onset of the pandemic and the corresponding lockdown in Germany. However, in a globally connected world, news from abroad could have easily affected search interests. In fact, search interest in COVID-19 reached its first peak before infection rates rose in Europe, especially when infection rates rose exponentially in China [32], suggesting a global influence [18]. The fact that regions with the highest numbers of cases were not always the first in terms of spreading information suggests a considerable role of cross-effects [19]. Consequently, regional, national, and even global events can exert an effect on public interest [31].

Some of the used search tags (eg, diarrhea) may not be uniquely related to the COVID-19 outbreak [11]. We addressed this challenge in two ways. First, we investigated relatively long baseline phases to test if interest after the arrival of the pandemic increased meaningfully compared to its baseline. Second, we did not base our analyses on single search tags that could easily be affected by third factors. Rather, we estimated interest in domains based on the aggregated information across several search tags.

Regression analysis revealed the necessity to account for correlated residuals during the linear modelling of Google Trends data. Another alternative for time series data would be the usage of vector autoregressive models (VAR) for prediction. VAR offers the advantage that criterion and predictor variables are interchangeable, allowing the use of all variables without prior assumptions about causality. This seems appropriate as cross-correlation functions in Multimedia Appendix 5 show moderate to high correlations between all studied time series. Even predicting new infection hot spots seems possible by analyzing public interest, given a proper sample size. Although this might be counterintuitive, there is supporting evidence that Google Trends data can be used to find new symptoms caused by diseases [41].

Although we z-standardized all variables prior to regression analysis, some betas exceeded the value of 1 (see Table 2, predictor NPI Social for the domains *vaccination* and *hobbies and sports*), possibly indicating suppression effects. However, post hoc correlation analysis between predictors indicated only moderate positive correlation. Thus, predictors are not highly collinear and betas exceeding 1 may reflect computational imprecision.

In line with our preregistration [49], our study was multivariate and multiple analyses had to be conducted. However, we did not correct statistical significances for multiple testing as we believe that effect sizes of the observed relations are more meaningful and closely in line with previous research for empirically derived domains. Nonetheless, we consider replication in independent samples very valuable to corroborate the obtained results, especially for the rationally selected domains used in the exploratory analyses.

Conclusions

This study shows that people's search interests are considerably affected by the pandemic and the issued NPIs. Increases were observed across a broad set of domains, reflecting psychological interests, motives, and concerns. For disease- and virus-associated domains, the increases occurred shortly after the onset of the pandemic, whereas domains capturing personal health strategies or community-level policies peaked later during the lockdown period. *Renovation* was the only domain to show an early decrease in public interest. The different onsets of the studied domains may indicate that some consequences are yet to become apparent, (eg, an increase in mental disorders due to novel life circumstances during the lockdown). Further research could, therefore, tap into multiple, more detailed domains using the demonstrated approach of TSFA while combining multiple keywords to handle undesirable error variance.

Using Google Trends data provides insights into people's search interests. The high sensitivity backs the interpretation of changes in public interest as a behavioral tendency [18,46]. Analyzing Google Trends data solves the problem of data scarcity, providing inexpensive information regarding possible behavioral changes that would be difficult to attain by means of questionnaire and survey data. This study shows pronounced changes in people's search interests during crises and lockdown, although prediction of specific developments is moderate at best. Given the relatively broad effects across the diverse domains, manifold effects on psychological interests, motives, and worries are indicated. In turn, these effects can help policy makers to develop a better insight into people's needs and concerns and to adapt and develop suitable information, prevention, and intervention strategies.

Acknowledgments

DR, PD, and FS contributed to conceptualization and methodology. DR performed data curation and formal analysis, as well as original draft preparation. DR, PD, and FS contributed to manuscript review and editing. We also thank Deniz Dericioğlu for his help in retrieving the keywords for our analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview of domains and corresponding search tags. Search tags we could not retrieve via Google Trends are marked with a *. Search tags that were dropped during data reduction and preprocessing are marked with a †.

[DOCX File , 43 KB - jmir_v23i6e26385_app1.docx]

Multimedia Appendix 2

Map showing the 16 German federal states colored by 7-day incidence of COVID-19 per 100,000 inhabitants for the middle (April 30, 2020) and the end of the lockdown (June 26, 2020) in Germany.

[DOCX File , 130 KB - jmir_v23i6e26385_app2.docx]

Multimedia Appendix 3

Summary statistics for population number and density, as well as cumulative infection and incidence rates for all federal states and Germany.

[DOCX File , 22 KB - jmir_v23i6e26385_app3.docx]

Multimedia Appendix 4

Scree plots for time series factor analysis for each domain after deleting items with negative loadings, in line with the preregistered protocol. All scree plots support the predicted underlying one-factor structure, indicating the communality of the respective domain.

[DOCX File , 332 KB - jmir_v23i6e26385_app4.docx]

Multimedia Appendix 5

Cross-correlation functions for each predictor for leads and lags of up to ±21 days for all domains used in the analysis. Predictive value as a function of lead/lag is given for infection rates, the countermeasure factors capturing regulation of outdoor activities (CM Outdoor) and regulation of social life (CM Social), and the truck toll mileage index as an indicator of the state of economy in Germany.

[DOCX File , 896 KB - jmir_v23i6e26385_app5.docx]

References

1. WHO announces COVID-19 outbreak a pandemic. World Health Organization. 2020 Mar 12. URL: https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic [accessed 2020-10-04]
2. Naming the coronavirus disease (COVID-19) and the virus that causes it. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it [accessed 2020-10-04]
3. Robert Koch-Institut: COVID-19-Dashboard. Robert Koch-Institut. URL: https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4 [accessed 2020-10-04]
4. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2020-10-04]
5. Steinmetz H, Batzdorfer V, Bosnjak M. The ZPID lockdown measures dataset. ZPID Sci Inf Online 2020;20(1):11. [doi: 10.23668/psycharchives.3019]
6. Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly C, Riley S, et al. Report 3: Transmissibility of 2019-nCoV. Imperial College London. 2020 Jan 25. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-transmissibility-25-01-2020.pdf [accessed 2021-05-19]
7. Bhatia S, Imai N, Cuomo-Dannenburg G, Baguelin M, Boonyasiri A, Cori A, et al. Estimating the number of undetected COVID-19 cases among travellers from mainland China. Wellcome Open Res 2020 Jun 15;5:143. [doi: 10.12688/wellcomeopenres.15805.1]
8. Ferguson N, Laydon D, Nedjati GG, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College London. 2020 Mar 16. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf [accessed 2021-05-21]
9. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Imperial College COVID-19 Response Team, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020 Aug 08;584(7820):257-261. [doi: 10.1038/s41586-020-2405-7] [Medline: 32512579]
10. Ainslie KEC, Walters CE, Fu H, Bhatia S, Wang H, Xi X, et al. Evidence of initial success for China exiting COVID-19 social distancing policy after achieving containment. Wellcome Open Res 2020 Oct 1;5:81 [FREE Full text] [doi: 10.12688/wellcomeopenres.15843.2] [Medline: 32500100]
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. 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]
13. 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]
14. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
15. Arora VS, McKee M, Stuckler D. Google Trends: Opportunities and limitations in health and health policy research. Health Policy 2019 Mar;123(3):338-341. [doi: 10.1016/j.healthpol.2019.01.001] [Medline: 30660346]
16. Kaleem T, Miller D, Waddle M, Stross W, Miller R, Trifiletti D. Google Search Trends in Oncology and the Impact of Celebrity Cancer Awareness. International Journal of Radiation Oncology, Biology, and Physics 2018 Nov;102(3):e270-e271. [doi: 10.1016/j.ijrobp.2018.07.876]
17. Ling R, Lee J. Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study. JMIR Public Health Surveill 2016 Oct 12;2(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]

18. Hu D, Lou X, Xu Z, Meng N, Xie Q, Zhang M, et al. More effective strategies are required to strengthen public awareness of COVID-19: Evidence from Google Trends. J Glob Health 2020 Jun;10(1):011003 [FREE Full text] [doi: 10.7189/jogh.10.011003] [Medline: 32373339]
19. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. JMIR Public Health Surveill 2020 May 05;6(2):e19374 [FREE Full text] [doi: 10.2196/19374] [Medline: 32338613]
20. Singh L, Bansal S, Bode L, Budak C, Chi G, Kawintiranon K, et al. A first look at COVID-19 information and misinformation sharing on Twitter. ArXiv. Preprint posted online on March 31, 2020 2020 Mar 31. [Medline: 32550244]
21. Rajan A, Sharaf R, Brown R, Sharaiha R, Lebwohl B, Mahadev S. Association of Search Query Interest in Gastrointestinal Symptoms With COVID-19 Diagnosis in the United States: Infodemiology Study. JMIR Public Health Surveill 2020 Jul 17;6(3):e19354 [FREE Full text] [doi: 10.2196/19354] [Medline: 32640418]
22. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
23. 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]
24. Rivers C, Chretien J, Riley S, Pavlin JA, Woodward A, Brett-Major D, et al. Using "outbreak science" to strengthen the use of models during epidemics. Nat Commun 2019 Jul 15;10(1):3102 [FREE Full text] [doi: 10.1038/s41467-019-11067-2] [Medline: 31308372]
25. Knipe D, Evans H, Marchant A, Gunnell D, John A. Mapping population mental health concerns related to COVID-19 and the consequences of physical distancing: a Google trends analysis. Wellcome Open Res 2020 Apr 30;5:82. [doi: 10.12688/wellcomeopenres.15870.1]
26. Heymann DL, Shindo N, WHO Scientific and Technical Advisory Group for Infectious Hazards. COVID-19: what is next for public health? Lancet 2020 Feb 22;395(10224):542-545 [FREE Full text] [doi: 10.1016/S0140-6736(20)30374-3] [Medline: 32061313]
27. Kardeş S, Kuzu AS, Pakhchanian H, Raiker R, Karagülle M. Population-level interest in anti-rheumatic drugs in the COVID-19 era: insights from Google Trends. Clin Rheumatol 2021 May;40(5):2047-2055 [FREE Full text] [doi: 10.1007/s10067-020-05490-w] [Medline: 33130946]
28. Kardeş S, Kuzu AS, Raiker R, Pakhchanian H, Karagülle M. Public interest in rheumatic diseases and rheumatologist in the United States during the COVID-19 pandemic: evidence from Google Trends. Rheumatol Int 2021 Feb;41(2):329-334 [FREE Full text] [doi: 10.1007/s00296-020-04728-9] [Medline: 33070255]
29. Strzelecki A. The second worldwide wave of interest in coronavirus since the COVID-19 outbreaks in South Korea, Italy and Iran: A Google Trends study. Brain Behav Immun 2020 Aug;88:950-951 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.042] [Medline: 32311493]
30. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad Sci U S A 2020 May 26;117(21):11220-11222 [FREE Full text] [doi: 10.1073/pnas.2005335117] [Medline: 32366658]
31. Husnayain A, Shim E, Fuad A, Su EC. Understanding the Community Risk Perceptions of the COVID-19 Outbreak in South Korea: Infodemiology Study. J Med Internet Res 2020 Sep 29;22(9):e19788 [FREE Full text] [doi: 10.2196/19788] [Medline: 32931446]
32. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends Analysis. Int J Infect Dis 2020 Jun;95:192-197 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.033] [Medline: 32305520]
33. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Jul;10(7):839-847 [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]
34. 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]
35. Cherry G, Rocke J, Chu M, Liu J, Lechner M, Lund VJ, et al. Loss of smell and taste: a new marker of COVID-19? Tracking reduced sense of smell during the coronavirus pandemic using search trends. Expert Rev Anti Infect Ther 2020 Nov;18(11):1165-1170 [FREE Full text] [doi: 10.1080/14787210.2020.1792289] [Medline: 32673122]
36. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):1-5 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
37. Yuan X, Xu J, Hussain S, Wang H, Gao N, Zhang L. Trends and Prediction in Daily New Cases and Deaths of COVID-19 in the United States: An Internet Search-Interest Based Model. Explor Res Hypothesis Med 2020 Apr 18;5(2):1-6 [FREE Full text] [doi: 10.14218/ERHM.2020.00023] [Medline: 32348380]

38. Lin Y, Chiang T, Lin Y. Increased Internet Searches for Insomnia as an Indicator of Global Mental Health During the COVID-19 Pandemic: Multinational Longitudinal Study. J Med Internet Res 2020 Sep 21;22(9):e22181 [FREE Full text] [doi: 10.2196/22181] [Medline: 32924951]
39. Mavragani A. Tracking COVID-19 in Europe: Infodemiology Approach. JMIR Public Health Surveill 2020 Apr 20;6(2):e18941 [FREE Full text] [doi: 10.2196/18941] [Medline: 32250957]
40. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
41. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
42. Brunori P, Resce G. Searching for the Peak Google Trends and the COVID-19 Outbreak in Italy. SSRN Journal 2020:1-13. [doi: 10.2139/ssrn.3569909]
43. Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis J, et al. Real-Time Forecasting of the COVID-19 Outbreak in Chinese Provinces: Machine Learning Approach Using Novel Digital Data and Estimates From Mechanistic Models. J Med Internet Res 2020 Aug 17;22(8):e20285 [FREE Full text] [doi: 10.2196/20285] [Medline: 32730217]
44. Cousins HC, Cousins CC, Harris A, Pasquale LR. Regional Infoveillance of COVID-19 Case Rates: Analysis of Search-Engine Query Patterns. J Med Internet Res 2020 Jul 30;22(7):e19483 [FREE Full text] [doi: 10.2196/19483] [Medline: 32692691]
45. Gane SB, Kelly C, Hopkins C. Isolated sudden onset anosmia in COVID-19 infection. A novel syndrome? Rhinology 2020 Jun 01;58(3):299-301. [doi: 10.4193/Rhin20.114] [Medline: 32240279]
46. Goldman D. Initial Observations of Psychological and Behavioral Effects of COVID-19 in the United States, Using Google Trends Data. SocArXiv 2020 Apr 04:1-5. [doi: 10.31235/osf.io/jecqp]
47. Brodeur A, Clark AE, Fleche S, Powdthavee N. COVID-19, lockdowns and well-being: Evidence from Google Trends. J Public Econ 2021 Jan;193:104346 [FREE Full text] [doi: 10.1016/j.jpubeco.2020.104346] [Medline: 33281237]
48. Schiller F, Bey K, Grabe L, Schmitz F. Stockpile purchasing in the emerging COVID-19 pandemic is related to obsessive-compulsiveness. Journal of Affective Disorders Reports 2021 Apr;4:100116. [doi: 10.1016/j.jadr.2021.100116]
49. Rotter D, Doebler P, Schmitz F. Information search in times of crisis: How Google trends can inform policy makers. OSF Registries. 2020. URL: https://osf.io/s5379 [accessed 2021-05-21]
50. Massicotte P, Eddelbuettel D. gtrendsR: Perform and Display Google Trends Queries. The Comprehensive R Archive Network. 2020. URL: https://cran.r-project.org/package=gtrendsR [accessed 2021-05-19]
51. R: A Language and Environment for Statistical Computing. R Core Team. 2017. URL: https://www.r-project.org/ [accessed 2021-05-19]
52. Situationsbericht vom 28.4.2020. Robert Koch Institut. 2020 Apr 28. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-04-28-de.pdf [accessed 2021-05-19]
53. Lkw-Maut-Fahrleistungsindex. Statistisches Bundesamt. URL: https://www.destatis.de/DE/Service/EXDAT/Datensaetze/lkw-maut-artikel.html [accessed 2020-10-04]
54. Gilbert P, Meijer E. Time series factor analysis with an application to measuring money. University of Groningen. 2005. URL: https://research.rug.nl/en/publications/time-series-factor-analysis-with-an-application-to-measuring-mone-2 [accessed 2021-05-21]
55. Rossiter D. Tutorial: Time series analysis in R. URL: http://www.css.cornell.edu/faculty/dgr2/_static/files/R_PDF/exTSA.pdf [accessed 2020-07-03]
56. Zeileis A, Leisch F, Hornik K, Kleiber C. strucchange: An R Package for Testing for Structural Change in Linear Regression Models. J Stat Soft 2002;7(2):1. [doi: 10.18637/jss.v007.i02]
57. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. Hoboken, NJ: John Wiley & Sons; 2015.
58. Schwarz G. Estimating the Dimension of a Model. Ann Statist 1978 Mar 1;6(2):461-464. [doi: 10.1214/aos/1176344136]
59. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Soft 2015;67(1):1. [doi: 10.18637/jss.v067.i01]
60. Share of households with internet access in Germany from 2007-2019. Statista GmbH. URL: https://www.statista.com/statistics/377677/household-internet-access-in-germany/ [accessed 2020-10-04]

Abbreviations

API: advanced programming interface
ARIMA: autoregressive integrated moving average
ICC: intraclass correlation
NPI: nonpharmaceutical intervention
NPI Social: NPI factor regulating social life

NPI Outdoor: NPI factor regulating outdoor activities
RKI: Robert Koch Institute
TSFA: time series factor analysis
TTMI: truck toll mileage index
VAR: vector autoregressive models
WHO: World Health Organization
ZPID: Leibniz Institute for Psychology

Edited by C Basch; submitted 09.12.20; peer-reviewed by A Mavragani, S Kardes; comments to author 13.01.21; revised version received 26.01.21; accepted 15.04.21; published 01.06.21.

Please cite as:
Rotter D, Doebler P, Schmitz F
Interests, Motives, and Psychological Burdens in Times of Crisis and Lockdown: Google Trends Analysis to Inform Policy Makers
J Med Internet Res 2021;23(6):e26385
URL: https://www.jmir.org/2021/6/e26385
doi:10.2196/26385
PMID:33999837

Original Paper

Topics of Nicotine-Related Discussions on Twitter: Infoveillance Study

Jon-Patrick Allem[1], MA, PhD; Allison Dormanesh[1], MS; Anuja Majmundar[2], PhD; Jennifer B Unger[1], PhD; Matthew G Kirkpatrick[1], PhD; Akshat Choube[3], MSc; Aneesh Aithal[3], MSc; Emilio Ferrara[3], PhD; Tess Boley Cruz[1], PhD

[1]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
[2]American Cancer Society, Washington, DC, United States
[3]Department of Computer Science, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Jon-Patrick Allem, MA, PhD
Department of Preventive Medicine
Keck School of Medicine
University of Southern California
2001 N Soto Street, 3rd Floor, SSB 312D
Los Angeles, CA, 90032
United States
Phone: 1 323 442 7921
Email: allem@usc.edu

Abstract

Background: Cultural trends in the United States, the nicotine consumer marketplace, and tobacco policies are changing.

Objective: The goal of this study was to identify and describe nicotine-related topics of conversation authored by the public and social bots on Twitter, including any misinformation or misconceptions that health education campaigns could potentially correct.

Methods: Twitter posts containing the term "nicotine" were obtained from September 30, 2018 to October 1, 2019. Methods were used to distinguish between posts from social bots and nonbots. Text classifiers were used to identify topics in posts (n=300,360).

Results: Prevalent topics of posts included vaping, smoking, addiction, withdrawal, nicotine health risks, and quit nicotine, with mentions of going "cold turkey" and needing help in quitting. Cessation was a common topic, with mentions of quitting and stopping smoking. Social bots discussed unsubstantiated health claims including how hypnotherapy, acupuncture, magnets worn on the ears, and time spent in the sauna can help in smoking cessation.

Conclusions: Health education efforts are needed to correct unsubstantiated health claims on Twitter and ultimately direct individuals who want to quit smoking to evidence-based cessation strategies. Future interventions could be designed to follow these topics of discussions on Twitter and engage with members of the public about evidence-based cessation methods in near real time when people are contemplating cessation.

(J Med Internet Res 2021;23(6):e25579) doi:10.2196/25579

KEYWORDS

nicotine; electronic cigarettes; Twitter; social media; social bots; cessation

Introduction

While combustible tobacco product use is declining in the United States, electronic cigarette (e-cigarette) use has risen in recent years among youth and young adults [1]. Nicotine is the primary psychoactive substance responsible for the abuse potential (ie, the likelihood that a substance will cause addiction) of combustible tobacco products and many e-cigarettes [2]. Like several other psychoactive drugs, including caffeine and amphetamines, nicotine produces acute central nervous system effects including increased heart rate, blood pressure, alertness, and decreased appetite [3], and both animal and human studies suggest that the drug may produce long-term deleterious effects on cognitive development among youth [3,4].

Research has repeatedly shown that there is substantial misunderstanding regarding the health risks of nicotine use [5].

While nicotine is the psychoactive component that sustains tobacco dependence [6], the primary carcinogenic harms are due to combustion of the tobacco leaf [3]. Nevertheless, one study demonstrated that 54% of smokers incorrectly believed that reductions in nicotine made cigarettes less dangerous [7]. Additionally, young adults (a priority population for tobacco control) commonly have misperceptions about the safety profile and nicotine content in e-cigarettes [8], including the unsubstantiated belief that e-cigarettes are relatively safe despite the burgeoning evidence indicating the products' nicotine-related abuse potential [9,10] and associations with progression to regular combustible cigarette use [11].

Availability of different e-cigarette products like those compatible with multiple substances (eg, open-system pod mods) [12-14] or products that facilitate customization may contribute to youth experimentation and transitions to combustible cigarette use. Such nicotine-use trajectories among youth make it crucial to characterize the public's experiences with, and perceptions of, nicotine.

Publicly accessible data from people who post to social media platforms, like Twitter, can be used to describe perceptions of nicotine and the social and environment context surrounding nicotine use [15]. Twitter is used by 22% of US adults (distributed fairly evenly through racial and gender groups), with 42% of users on the platform daily [16]. Twitter is also used by 32% of adolescents (13 to 17 years old) in the United States [17]. Previous analyses of posts to Twitter have provided insight about what the public organically discusses regarding tobacco, including the frequency of use, co-use with other substances (eg, alcohol, marijuana), mentions of tobacco product appeal, and the locations where tobacco is often used [18,19]. Past literature also highlights the role of social bots (ie, automated accounts created to produce content and interact with human accounts on Twitter) in spreading unsubstantiated health claims and misinformation on health-related topics such as vaping and vaccines [20,21]. The goal of this study was to identify and describe nicotine-related topics of conversation authored by the public and social bots on Twitter, including any misinformation or misconceptions that health education campaigns could potentially correct.

Methods

Twitter posts containing the term "nicotine" ("#nicotine" would also be included in this search) were obtained from Twitter's Streaming Application Program Interface (API; the filtered stream using the Twitter4J library for collecting tweets with no gaps in the collection time) from September 30, 2018 to October 1, 2019. There was a total of 1,203,466 posts containing this term during this time. Similar to prior research [15,18], we removed all retweets (n=786,327) and non-English tweets (n=45,497), resulting in 371,642 unique tweets. Removing retweets allowed us to treat each observation as independent. Posts that contained the term "nicotine" but were determined to be unrelated to our research objectives were identified and removed. This included tweets containing the phrases, "bad nicotine," "nicotine heroin," "nicotine stain," and "silver spoon," as these were references to popular song lyrics. As a result of this filtering process, we were left with 364,430 unique tweets.

Next, we identified social bots [20]. Social bots may bias the data, reducing our ability to dependably describe the public's recent experience with nicotine [22]. We used Botometer [23] to distinguish between nonbots and social bots. Botometer analyzes the characteristics of a Twitter account and scores it based on how likely the account is to be a social bot. It is considered a state-of-the-art machine learning algorithm and has been used in prior research revolving around social bots and public health [15,21,24]. The Botometer threshold was set to ≥4 on the scale out of 5 of English scores and similar to prior research [25]. Each Twitter account was screened after posts were collected (ie, not in real time). During this process, Twitter accounts (n=27,186) responsible for posts in our data had been deleted. Because these Twitter accounts ceased to exist and could not be processed through Botometer, we removed the posts (n=42,890) from these accounts from our data. The final sample contained 321,540 posts, with 300,360 posts from 181,439 unique nonbot accounts, and 21,180 posts from 5889 social bots.

All analyses relied on public, anonymized data; adhered to the terms and conditions, terms of use, and privacy policies of Twitter; and were performed under the institutional review board approval from the authors' university. To protect privacy, no tweets were reported verbatim in this article. To promote full transparency and foster reproducibility, all data and code are available from the lead author and posted on his website and data repository.

To prepare tweets for analysis, we conducted a number of transformations, including (1) basic normalization (ie, lower casing all tweets; removing extra spaces, punctuation, and special characters such as brackets), (2) stop word removal (ie, removing words such as "a," "the"), (3) normalizing Twitter account mentions (ie, @account_name occurrences in the tweets were replaced by @person — a common token for all accounts), (4) lemmatization (ie, the removal of inflections and variants of words), (5) nonprintable character removal (ie, removing emoticons or as symbols from non-English languages), and (6) removal of hashtags and URLs.

To find topics within the tweets, we generated n-grams for n=1 (ie, unigrams) and n=2 (ie, bigrams) from each tweet. An n-gram is simply a sequence of n words. For example, the phrase "Player breaks record" contains the unigrams "player," "breaks," "record" and the bigrams "player breaks" and "breaks record." By generating frequency counts of the most common unigrams and bigrams, we obtained an initial sense of the commonly discussed topics. From this assessment of the most common words and phrases, 4 of the authors reviewed posts in their entirety and arrived at a consensus on 15 commonly occurring topics. This strategy was used to summarize the raw text-based data, documenting the patterns that were present. Topics included person tagging (@person), addiction (mentions of being addicted to nicotine or craving nicotine), appeal (mentions of liking or loving nicotine), nicotine replacement therapies (NRT; mentions of the patch, gum, nicotine replacement), vaping (mentions of using e-cigarettes, vaping, JUUL), smoking (mentions of smoking cigarettes, using other combustible tobacco), nicotine health risks (mentions of nicotine effects on the brain, respiratory health, the amount of nicotine in products), withdrawal (mentions of nicotine withdrawal), quit nicotine (mentions of quitting nicotine or going nicotine free), cessation (mentions of quitting or stopping smoking), polysubstance use (mentions of alcohol and nicotine use), caffeine (mentions of coffee and nicotine use), underage use (mentions of children and teens using nicotine, use of nicotine at high schools), and new products (mentions of a "nicotine shot" or a supplement to boost the amount of nicotine in e-liquids). Nicotine is safe (mentions of nicotine not being harmful by itself) was a topic established a priori since these posts may reflect misconceptions that could be addressed by health education campaigns [26].

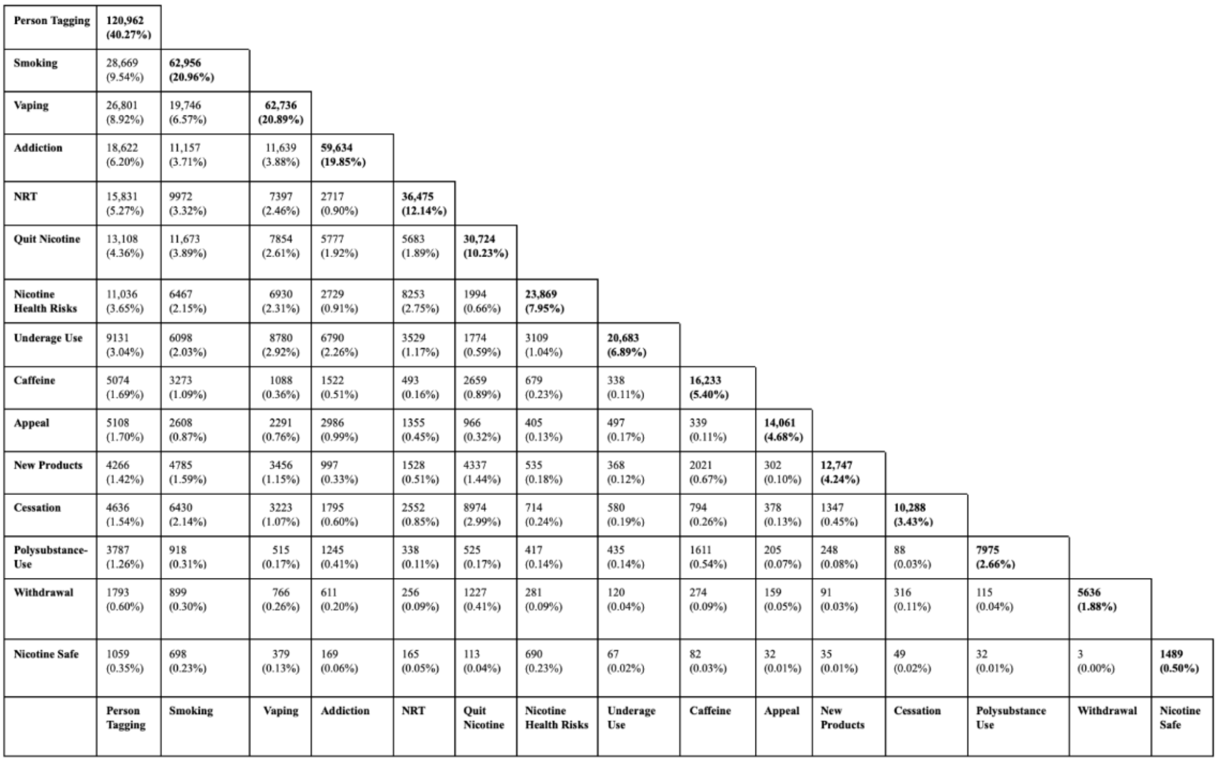
Each tweet was classified to one or more topics based on the occurrence of at least one topic-related pattern, which is similar to prior research [18,25]. This pattern could be a unigram, a bigram, or groups of words that must occur in the normalized tweets in a specific order. This was accomplished by using a rule-based classification algorithm developed in Python that inspects each tweet for the presence of a specified set of patterns representing a topic. Since a single post could discuss multiple topics, we report the percentage of overlap between each topic by utilizing a confusion matrix. Each cell in the matrix represents the intersection of 2 topics. The value of the cell represents the percentage of the total corpus that belongs to both topics. For example, a hypothetical post such as "Hey @person look who is nicotine free today" would be classified under "person tagging" and "quit nicotine." The number of posts containing both would be found at the intersection of the matrix for these 2 topics.

Results

The total coverage of the 15 topics constituted 82.86% (248,893/300,360) of all tweets in the corpus from nonbots (Figure 1). The remaining 17.14% (51,467/300,360) of tweets were too diverse to be classified into a single topic with meaningful coverage (ie, coverage of each subsequent topic would be less than 1% of total tweets). The most prevalent topic in this corpus was "person tagging" at 40.27% (120,962/300,360), followed by "smoking" at 20.96% (62,956/300,360) and "vaping" at 20.89% (62,736/300,360). "Addiction" was the next most prevalent topic at 19.85% (59,634/300,360), followed by "NRT" at 12.14 % (36,475/300,360) and "quit nicotine" at 10.23% (30,724/300,360). Among "quit nicotine," posts suggested going "cold turkey," a day without nicotine, trying to quit, and needing help in quitting. "Nicotine health risks" was a common topic at 7.95% (23,869/300,360), followed by "underage use" at 6.89% (20,683/300,360), "caffeine" at 5.40% (16,233/300,360), "appeal" at 4.68% (14,061/300,360), "new products" at 4.24% (12,747/300,360), and "cessation" at 3.43% (10,288/300,360). Among "cessation," posts suggested quitting and stopping smoking. "Nicotine is safe" was an uncommon topic at 0.50% (1489/300,360).

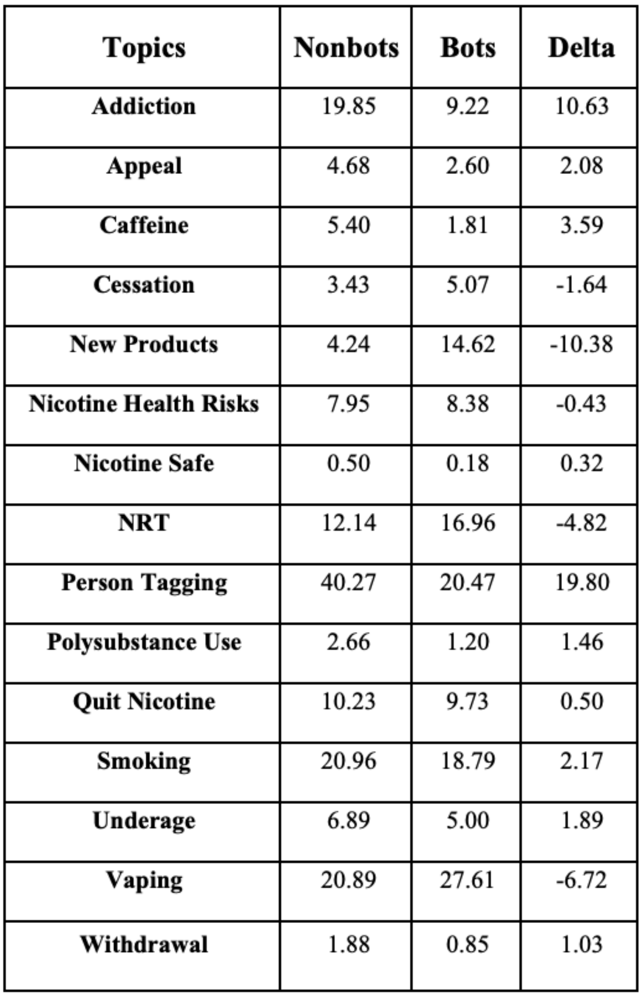
The total coverage of the same 15 topics constituted 75.56% (16,004/21,180) of all tweets in the corpus from social bots (Figure 2). Comparing the 2 corpora, some topics had similar prevalences, while other topics stood out with large differences. For example, the largest difference in prevalence in topics between corpora was found in "person tagging" (nonbots at 40.27% [120,962/300,360] versus social bots at 20.47% [4336/21,180]), followed by "new products" (nonbots at 4.24% [12,747/300,360] versus social bots 14.62% [3096/21,180]) and "addiction" (nonbots at 19.85% [59,634/300,360] versus social bots at 9.22% [1952/21,180]). The content found in each category was overall consistent between nonbots and social bots in all but "cessation." Posts in "cessation" from social bots regularly included the use of hypnotherapy, acupuncture, magnets worn on the ears, and time spent in the sauna as effective ways to stop smoking.

Figure 1. Prevalence of topics from nonbot corpus. NRT: nicotine replacement therapy.



Person Tagging	**120,962 (40.27%)**														
Smoking	28,669 (9.54%)	**62,956 (20.96%)**													
Vaping	26,801 (8.92%)	19,746 (6.57%)	**62,736 (20.89%)**												
Addiction	18,622 (6.20%)	11,157 (3.71%)	11,639 (3.88%)	**59,634 (19.85%)**											
NRT	15,831 (5.27%)	9972 (3.32%)	7397 (2.46%)	2717 (0.90%)	**36,475 (12.14%)**										
Quit Nicotine	13,108 (4.36%)	11,673 (3.89%)	7854 (2.61%)	5777 (1.92%)	5683 (1.89%)	**30,724 (10.23%)**									
Nicotine Health Risks	11,036 (3.65%)	6467 (2.15%)	6930 (2.31%)	2729 (0.91%)	8253 (2.75%)	1994 (0.66%)	**23,869 (7.95%)**								
Underage Use	9131 (3.04%)	6098 (2.03%)	8780 (2.92%)	6790 (2.26%)	3529 (1.17%)	1774 (0.59%)	3109 (1.04%)	**20,683 (6.89%)**							
Caffeine	5074 (1.69%)	3273 (1.09%)	1088 (0.36%)	1522 (0.51%)	493 (0.16%)	2659 (0.89%)	679 (0.23%)	338 (0.11%)	**16,233 (5.40%)**						
Appeal	5108 (1.70%)	2608 (0.87%)	2291 (0.76%)	2986 (0.99%)	1355 (0.45%)	966 (0.32%)	405 (0.13%)	497 (0.17%)	339 (0.11%)	**14,061 (4.68%)**					
New Products	4266 (1.42%)	4785 (1.59%)	3456 (1.15%)	997 (0.33%)	1528 (0.51%)	4337 (1.44%)	535 (0.18%)	368 (0.12%)	2021 (0.67%)	302 (0.10%)	**12,747 (4.24%)**				
Cessation	4636 (1.54%)	6430 (2.14%)	3223 (1.07%)	1795 (0.60%)	2552 (0.85%)	8974 (2.99%)	714 (0.24%)	580 (0.19%)	794 (0.26%)	378 (0.13%)	1347 (0.45%)	**10,288 (3.43%)**			
Polysubstance-Use	3787 (1.26%)	918 (0.31%)	515 (0.17%)	1245 (0.41%)	338 (0.11%)	525 (0.17%)	417 (0.14%)	435 (0.14%)	1611 (0.54%)	205 (0.07%)	248 (0.08%)	88 (0.03%)	**7975 (2.66%)**		
Withdrawal	1793 (0.60%)	899 (0.30%)	766 (0.26%)	611 (0.20%)	256 (0.09%)	1227 (0.41%)	281 (0.09%)	120 (0.04%)	274 (0.09%)	159 (0.05%)	91 (0.03%)	316 (0.11%)	115 (0.04%)	**5636 (1.88%)**	
Nicotine Safe	1059 (0.35%)	698 (0.23%)	379 (0.13%)	169 (0.06%)	165 (0.05%)	113 (0.04%)	690 (0.23%)	67 (0.02%)	82 (0.03%)	32 (0.01%)	35 (0.01%)	49 (0.02%)	32 (0.01%)	3 (0.00%)	**1489 (0.50%)**
	Person Tagging	**Smoking**	**Vaping**	**Addiction**	**NRT**	**Quit Nicotine**	**Nicotine Health Risks**	**Underage Use**	**Caffeine**	**Appeal**	**New Products**	**Cessation**	**Polysubstance Use**	**Withdrawal**	**Nicotine Safe**

Figure 2. Comparison of prevalence of topics between nonbots and social bots. NRT: nicotine replacement therapy.



Topics	Nonbots	Bots	Delta
Addiction	19.85	9.22	10.63
Appeal	4.68	2.60	2.08
Caffeine	5.40	1.81	3.59
Cessation	3.43	5.07	-1.64
New Products	4.24	14.62	-10.38
Nicotine Health Risks	7.95	8.38	-0.43
Nicotine Safe	0.50	0.18	0.32
NRT	12.14	16.96	-4.82
Person Tagging	40.27	20.47	19.80
Polysubstance Use	2.66	1.20	1.46
Quit Nicotine	10.23	9.73	0.50
Smoking	20.96	18.79	2.17
Underage	6.89	5.00	1.89
Vaping	20.89	27.61	-6.72
Withdrawal	1.88	0.85	1.03

Discussion

Principal Findings

This study is one of the largest Twitter studies to date focused on nicotine-related conversations, describing over 300,000 unique posts from over 180,000 unique accounts and addressing the underlying questions of what the public discusses or perceives about nicotine (rather than focusing on one specific tobacco product). We identified a number of topics of conversation ranging from nicotine appeal to withdrawal to smoking cessation. Posts discussed addiction, NRT, health risks, and nicotine use in combination with alcohol and caffeine. This study also distinguished nicotine-related topics of conversations by social bots and nonbots, describing differences in prevalence of topics by account type.

In this study, Twitter posts mentioning new products represented a larger proportion of posts by social bots compared to nonbots, suggesting that companies or retailers or e-cigarette hobbyists may be using bots to promote new products. Social bots have previously been found to promote emerging products on Twitter; for example, in 2017, it was found that social bots were more than 2 times as likely to post about a new vaping product compared to nonbots [15]. Posts from social bots identified in the present study perpetuated a number of methods with very limited evidence as smoking cessation interventions, including hypnotherapy, acupuncture, trips to the sauna, and the use of magnets behind the ear. In contrast to front-line treatments such as tailored behavioral counseling (eg, individual, group, and phone) and medication (eg, varenicline, bupropion, NRT), these alternative methods have little to no empirical evidence to support their efficacy [27,28]. Unsubstantiated health claims perpetuated by social bots may have offline consequences, such as leaving Twitter users with the impression that these methods are good cessation strategies, thus diverting them from more effective approaches.

Unsubstantiated health claims on Twitter from social bots have been documented in prior research; for example, several studies have reported that social bots regularly make claims touting the effectiveness of e-cigarettes in smoking cessation [15,24] and claims propagating misinformation pertaining to vaccinations [21]. Recently, it was reported that social bots were responsible for disseminating unsubstantiated health claims pertaining to cannabis with posts suggesting cannabis could allay health concerns ranging from triple-negative breast cancer to plantar fasciitis [25]. Health education efforts are needed to correct misinformation and ultimately direct individuals who want to quit smoking to evidence-based cessation strategies [27,29]. Misperceptions or myths about cessation could be most persuasively countered with two-sided messages that provide a brief acknowledgement of the misconception, then refute it, and followed by a stronger statement about the more effective intervention [30]. For example, Twitter posts could be circulated that state: "If you feel addicted to cigarettes, you could try quitting cold turkey or with hypnotherapy, but you are more likely to succeed if you work with a Quitline like 1-800-NO BUTTS."

"Person tagging" was a predominant theme in the current study of nicotine-related posts to Twitter and in line with prior research [18,25]. Person tagging in this context is a social practice where Twitter users directly interact with one another to exchange their attitudes about and experiences with nicotine. Posts classified under "person tagging" regularly used @Person to engage others in discussions about nicotine. These online communications may impact nicotine use; for example, Unger and colleagues [31] demonstrated an empirical link between adolescents' and young adults' tobacco-related Twitter activity and their tobacco product use. The current study's findings are highly relevant to the public health community, as repeated exposure to nicotine-related messaging and reported nicotine use by Twitter connections may influence the social norms of those exposed to the content and lead to imitation of the behaviors [32].

Prior research has shown that a cessation program utilizing Twitter to deliver an intervention for smoking cessation can be successful in helping participants sustain abstinence [33]. The present study did not identify participants looking to quit smoking on Twitter; however, these findings suggest that Twitter may be a place where such participants could be found as people tweet about the difficulty of quitting nicotine. "Vaping," "addiction," "quit nicotine," "withdrawal," and "cessation" were all topics in the present study. Future interventions could be designed to follow these topics of discussions on Twitter and engage with potential participants about evidence-based cessation methods in near real time when people are contemplating cessation [34].

"Polysubstance use" and "caffeine" were identified as topics in the current study. Polysubstance use has been reported in several earlier Twitter-based studies; for example, a prior analysis of hookah-related posts to Twitter from 2017 to 2018 found that many posts described alcohol, marijuana, and other substance use along with hookah [18]. Similar findings were reported in Twitter studies focused on e-cigarettes [15] and cannabis [35]. Past work also raises concerns about the unknown health effects of caffeine in flavored e-liquids [36] and preference of e-liquids with active caffeine ingredients for weight loss [37]. The present findings supplement these previous studies and further awareness of the occurrence of polysubstance use. This is particularly important because alcohol and caffeine can potentiate the reinforcing effects of nicotine [38,39], potentially leading to escalation in use of one or both substances.

Similar to prior Twitter studies focused on JUUL use [40], the current study found posts indicative of underage use of nicotine (ie, mentions of nicotine use at high schools and among teenagers). This finding is concerning because nicotine impairs adolescents' and young adults' brain development [2,3,41]. In addition, posts about underage use may normalize e-cigarettes in young viewers, with the potential to increase experimentation and regular use [42].

Limitations

This study focused on posts to Twitter, and findings may not extend to other social media platforms. The posts in this study were collected within a 12-month period and may not extend to other time periods. Data collection relied on Twitter's Streaming API, which prevented collection of posts from private accounts. Findings may not generalize to all Twitter users or to the US population. Not all tweets were covered by the established categories, and topics of conversation were not segmented by geographic location, preventing this study from understanding the effect of different state tobacco policies on the public's experience with nicotine. Prior research has shown that significant geographic biases can occur in the context of conversations over Twitter [43,44]. In some instances, unigrams and bigrams used to define topics may have multiple meanings that were ignored in the current study; for example, the word "school" in nicotine-related posts may not always indicate underage use, as college students or other educational professionals may be discussing nicotine use.

Conclusions

Common nicotine-related topics on Twitter included smoking, vaping, cessation, withdrawal, and appeal, among others. These results suggest that Twitter users often discuss grappling with quitting smoking, nicotine withdrawal, and nicotine cravings. Such topics of conversation warrant considerations by public health researchers in the future. Twitter may act as a platform to engage with those struggling with nicotine dependence, as well as those initiating use with nicotine-related products, by informing them of the potential for dependence and subsequent health consequences of use. Posts from social bots regularly included the use of hypnotherapy, acupuncture, magnets worn on the ears, and time spent in the sauna as effective ways to stop smoking. Misinformation regarding nicotine has been a component of tobacco industry marketing and has the potential to influence beliefs, perceptions, and use of tobacco; thus, it is important to provide a recent account of what posts discuss on Twitter about nicotine in hopes of correcting misinformation and directing tobacco users to more effective interventions.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute and the Food and Drug Administration Center for Tobacco Products (CTP) [Grant # U54 CA180905]. The funders had no role in study design; collection, analysis, and interpretation of data; writing the report; and the decision to submit the report for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

Conflicts of Interest

None declared.

References

1. Dai H, Leventhal AM. Prevalence of e-Cigarette Use Among Adults in the United States, 2014-2018. JAMA 2019 Nov 12;322(18):1824-1827 [FREE Full text] [doi: 10.1001/jama.2019.15331] [Medline: 31524940]
2. England LJ, Bunnell RE, Pechacek TF, Tong VT, McAfee TA. Nicotine and the Developing Human: A Neglected Element in the Electronic Cigarette Debate. Am J Prev Med 2015 Aug;49(2):286-293 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.015] [Medline: 25794473]
3. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention (US); 2014.
4. Jacobsen LK, Krystal JH, Mencl WE, Westerveld M, Frost SJ, Pugh KR. Effects of smoking and smoking abstinence on cognition in adolescent tobacco smokers. Biol Psychiatry 2005 Jan 01;57(1):56-66. [doi: 10.1016/j.biopsych.2004.10.022] [Medline: 15607301]
5. Czoli CD, Fong GT, Mays D, Hammond D. How do consumers perceive differences in risk across nicotine products? A review of relative risk perceptions across smokeless tobacco, e-cigarettes, nicotine replacement therapy and combustible cigarettes. Tob Control 2017 Mar 13;26(e1):e49-e58. [doi: 10.1136/tobaccocontrol-2016-053060] [Medline: 27625408]
6. Benowitz NL. Pharmacology of nicotine: addiction, smoking-induced disease, and therapeutics. Annu Rev Pharmacol Toxicol 2009;49:57-71 [FREE Full text] [doi: 10.1146/annurev.pharmtox.48.113006.094742] [Medline: 18834313]
7. Cummings KM, Hyland A, Giovino G, Hastrup J, Bauer J, Bansal M. Are smokers adequately informed about the health risks of smoking and medicinal nicotine? Nicotine Tob Res 2004 Dec;6 Suppl 3(6):S333-S340. [doi: 10.1080/14622200412331320734] [Medline: 15799596]
8. Fadus MC, Smith TT, Squeglia LM. The rise of e-cigarettes, pod mod devices, and JUUL among youth: Factors influencing use, health implications, and downstream effects. Drug Alcohol Depend 2019 Aug 01;201:85-93 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.04.011] [Medline: 31200279]
9. Maloney S, Breland A, Soule E, Hiler M, Ramôa C, Lipato T, et al. Abuse liability assessment of an electronic cigarette in combustible cigarette smokers. Exp Clin Psychopharmacol 2019 Oct;27(5):443-454 [FREE Full text] [doi: 10.1037/pha0000261] [Medline: 30777773]
10. Cobb CO, Lopez AA, Soule EK, Yen M, Rumsey H, Lester Scholtes R, et al. Influence of electronic cigarette liquid flavors and nicotine concentration on subjective measures of abuse liability in young adult cigarette smokers. Drug Alcohol Depend 2019 Oct 01;203:27-34. [doi: 10.1016/j.drugalcdep.2019.05.024] [Medline: 31401532]
11. Barrington-Trimis JL, Kong G, Leventhal AM, Liu F, Mayer M, Cruz TB, et al. E-cigarette Use and Subsequent Smoking Frequency Among Adolescents. Pediatrics 2018 Dec;142(6):1 [FREE Full text] [doi: 10.1542/peds.2018-0486] [Medline: 30397165]
12. Chaffee BW, Watkins SL, Glantz SA. Electronic Cigarette Use and Progression From Experimentation to Established Smoking. Pediatrics 2018 Sep 31;142(3):e20181885 [FREE Full text] [doi: 10.1542/peds.2018-1885] [Medline: 30177516]
13. Stanton C, Bansal-Travers M, Johnson A, Sharma E, Katz L, Ambrose B, et al. Longitudinal e-Cigarette and Cigarette Use Among US Youth in the PATH Study (2013-2015). J Natl Cancer Inst 2019 Oct 01;111(10):1088-1096 [FREE Full text] [doi: 10.1093/jnci/djz006] [Medline: 30689915]
14. Majmundar A, Kirkpatrick M, Cruz TB, Unger JB, Allem J. Characterising KandyPens-related posts to Instagram: implications for nicotine and cannabis use. Tob Control 2020 Jul 30;29(4):472-474. [doi: 10.1136/tobaccocontrol-2019-055006] [Medline: 31147484]
15. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-Cigarette Surveillance With Social Media Data: Social Bots, Emerging Topics, and Trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
16. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. 2019 Apr 10. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2021-05-18]
17. Anderson M, Jiang J. Pew Research Center. 2018 May 31. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2021-05-18]
18. Allem J, Dharmapuri L, Leventhal A, Unger J, Boley Cruz T. Hookah-Related Posts to Twitter From 2017 to 2018: Thematic Analysis. J Med Internet Res 2018 Nov 19;20(11):e11669 [FREE Full text] [doi: 10.2196/11669] [Medline: 30455162]
19. Majmundar A, Allem J, Cruz TB, Unger JB. Where Do People Vape? Insights from Twitter Data. Int J Environ Res Public Health 2019 Aug 23;16(17):3056 [FREE Full text] [doi: 10.3390/ijerph16173056] [Medline: 31443591]
20. Allem J, Ferrara E. Could Social Bots Pose a Threat to Public Health? Am J Public Health 2018 Aug;108(8):1005-1006. [doi: 10.2105/ajph.2018.304512]

21. Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, et al. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. Am J Public Health 2018 Oct;108(10):1378-1384. [doi: 10.2105/AJPH.2018.304567] [Medline: 30138075]
22. Allem J, Ferrara E. The Importance of Debiasing Social Media Data to Better Understand E-Cigarette-Related Attitudes and Behaviors. J Med Internet Res 2016 Aug 09;18(8):e219 [FREE Full text] [doi: 10.2196/jmir.6185] [Medline: 27507563]
23. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. Commun. ACM 2016 Jun 24;59(7):96-104. [doi: 10.1145/2818717]
24. Martinez LS, Hughes S, Walsh-Buhi ER, Tsou M. "Okay, We Get It. You Vape": An Analysis of Geocoded Content, Context, and Sentiment regarding E-Cigarettes on Twitter. J Health Commun 2018 Jul 06;23(6):550-562. [doi: 10.1080/10810730.2018.1493057] [Medline: 29979920]
25. Allem J, Escobedo P, Dharmapuri L. Cannabis Surveillance With Twitter Data: Emerging Topics and Social Bots. Am J Public Health 2020 Mar;110(3):357-362. [doi: 10.2105/ajph.2019.305461]
26. Huang J, Feng B, Weaver SR, Pechacek TF, Slovic P, Eriksen MP. Changing Perceptions of Harm of e-Cigarette vs Cigarette Use Among Adults in 2 US National Surveys From 2012 to 2017. JAMA Netw Open 2019 Mar 01;2(3):e191047 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.1047] [Medline: 30924893]
27. 2008 PHS Guideline Update Panel, Liaisons, and Staff. Treating tobacco use and dependence: 2008 update U.S. Public Health Service Clinical Practice Guideline executive summary. Respir Care 2008 Sep;53(9):1217-1222 [FREE Full text] [Medline: 18807274]
28. Barnes J, McRobbie H, Dong CY, Walker N, Hartmann-Boyce J. Hypnotherapy for smoking cessation. Cochrane Database Syst Rev 2019 Jun 14;6:CD001008 [FREE Full text] [doi: 10.1002/14651858.CD001008.pub3] [Medline: 31198991]
29. Final Recommendation Statement: Interventions for Tobacco Smoking Cessation in Adults, Including Pregnant Persons. US Preventive Services Task Force. 2021 Jan 19. URL: https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/tobacco-use-in-adults-and-pregnant-women-counseling-and-interventions1 [accessed 2021-05-18]
30. Allen M. Meta‐analysis comparing the persuasiveness of one‐sided and two‐sided messages. Western Journal of Speech Communication 1991 Dec 30;55(4):390-404. [doi: 10.1080/10570319109374395]
31. Unger JB, Urman R, Cruz TB, Majmundar A, Barrington-Trimis J, Pentz MA, et al. Talking about tobacco on Twitter is associated with tobacco product use. Prev Med 2018 Sep;114:54-56 [FREE Full text] [doi: 10.1016/j.ypmed.2018.06.006] [Medline: 29898418]
32. Fujimoto K, Valente TW. Social network influences on adolescent substance use: disentangling structural equivalence from cohesion. Soc Sci Med 2012 Jun;74(12):1952-1960 [FREE Full text] [doi: 10.1016/j.socscimed.2012.02.009] [Medline: 22475405]
33. Pechmann C, Delucchi K, Lakon CM, Prochaska JJ. Randomised controlled trial evaluation of Tweet2Quit: a social network quit-smoking intervention. Tob Control 2017 Mar 29;26(2):188-194 [FREE Full text] [doi: 10.1136/tobaccocontrol-2015-052768] [Medline: 26928205]
34. Deb A, Majmundar A, Seo S, Matsui A, Tandon R, Yan S, et al. Social Bots for Online Public Health Interventions. 2018 Presented at: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); August 28-31, 2018; Barcelona, Spain. [doi: 10.1109/asonam.2018.8508382]
35. Krauss MJ, Grucza RA, Bierut LJ, Cavazos-Rehg PA. “Get drunk. Smoke weed. Have fun.”: A Content Analysis of Tweets About Marijuana and Alcohol. Am J Health Promot 2016 Nov 17;31(3):200-208. [doi: 10.4278/ajhp.150205-qual-708]
36. Lisko JG, Lee GE, Kimbrell JB, Rybak ME, Valentin-Blasini L, Watson CH. Caffeine Concentrations in Coffee, Tea, Chocolate, and Energy Drink Flavored E-liquids. Nicotine Tob Res 2017 Apr 01;19(4):484-492 [FREE Full text] [doi: 10.1093/ntr/ntw192] [Medline: 27613945]
37. Morean ME, Wedel AV. Vaping to lose weight: Predictors of adult e-cigarette use for weight loss or control. Addict Behav 2017 Mar;66:55-59. [doi: 10.1016/j.addbeh.2016.10.022] [Medline: 27875790]
38. Jones HE, Griffiths RR. Oral caffeine maintenance potentiates the reinforcing and stimulant subjective effects of intravenous nicotine in cigarette smokers. Psychopharmacology (Berl) 2003 Jan;165(3):280-290. [doi: 10.1007/s00213-002-1262-4] [Medline: 12434259]
39. Rose JE, Brauer LH, Behm FM, Cramblett M, Calkins K, Lawhon D. Potentiation of Nicotine Reward by Alcohol. Alcoholism Clin Exp Res 2002 Dec;26(12):1930-1931. [doi: 10.1111/j.1530-0277.2002.tb02507.x]
40. Allem J, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related posts on Twitter. Drug Alcohol Depend 2018 Sep 01;190:1-5 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.05.018] [Medline: 29958115]
41. Jasinska AJ, Zorick T, Brody AL, Stein EA. Dual role of nicotine in addiction and cognition: a review of neuroimaging studies in humans. Neuropharmacology 2014 Sep;84:111-122 [FREE Full text] [doi: 10.1016/j.neuropharm.2013.02.015] [Medline: 23474015]
42. Gorukanti A, Delucchi K, Ling P, Fisher-Travis R, Halpern-Felsher B. Adolescents' attitudes towards e-cigarette ingredients, safety, addictive properties, social norms, and regulation. Prev Med 2017 Jan;94:65-71 [FREE Full text] [doi: 10.1016/j.ypmed.2016.10.019] [Medline: 27773711]

43. 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]
44. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018 Jun 14;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]

Abbreviations

API: application program interface
NRT: nicotine replacement therapy

Edited by R Kukafka; submitted 06.11.20; peer-reviewed by A Weke, J Brigham, R Gore; comments to author 26.11.20; revised version received 08.12.20; accepted 13.05.21; published 07.06.21.

Please cite as:
Allem JP, Dormanesh A, Majmundar A, Unger JB, Kirkpatrick MG, Choube A, Aithal A, Ferrara E, Boley Cruz T
Topics of Nicotine-Related Discussions on Twitter: Infoveillance Study
J Med Internet Res 2021;23(6):e25579
URL: https://www.jmir.org/2021/6/e25579
doi:10.2196/25579
PMID:34096875

Original Paper

A Determinants-of-Fertility Ontology for Detecting Future Signals of Fertility Issues From Social Media Data: Development of an Ontology

Ji-Hyun Lee[1], RN, PhD; Hyeoun-Ae Park[2], RN, PhD; Tae-Min Song[3], PhD

[1]Nurse's Office, Yeongdeok High School, Gyeonggi-do, Republic of Korea
[2]College of Nursing and Research Institute of Nursing Science, Seoul National University, Seoul, Republic of Korea
[3]Department of Health Management, Samyook University, Seoul, Republic of Korea

Corresponding Author:
Hyeoun-Ae Park, RN, PhD
College of Nursing and Research Institute of Nursing Science
Seoul National University
103 Daehak-ro, Jongno-gu
Seoul, 03080
Republic of Korea
Phone: 82 27408827
Fax: 82 27661852
Email: hapark@snu.ac.kr

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/7/e31601

Abstract

Background: South Korea has the lowest fertility rate in the world despite considerable governmental efforts to boost it. Increasing the fertility rate and achieving the desired outcomes of any implemented policies requires reliable data on the ongoing trends in fertility and preparations for the future based on these trends.

Objective: The aims of this study were to (1) develop a determinants-of-fertility ontology with terminology for collecting and analyzing social media data; (2) determine the description logics, content coverage, and structural and representational layers of the ontology; and (3) use the ontology to detect future signals of fertility issues.

Methods: An ontology was developed using the Ontology Development 101 methodology. The domain and scope of the ontology were defined by compiling a list of competency questions. The terms were collected from Korean government reports, Korea's Basic Plan for Low Fertility and Aging Society, a national survey about marriage and childbirth, and social media postings on fertility issues. The classes and their hierarchy were defined using a top-down approach based on an ecological model. The internal structure of classes was defined using the entity-attribute-value model. The description logics of the ontology were evaluated using Protégé (version 5.5.0), and the content coverage was evaluated by comparing concepts extracted from social media posts with the list of ontology classes. The structural and representational layers of the ontology were evaluated by experts. Social media data were collected from 183 online channels between January 1, 2011, and June 30, 2015. To detect future signals of fertility issues, 2 classes of the ontology, the socioeconomic and cultural environment, and public policy, were identified as keywords. A keyword issue map was constructed, and the defined keywords were mapped to identify future signals. R software (version 3.5.2) was used to mine for future signals.

Results: A determinants-of-fertility ontology comprised 236 classes and terminology comprised 1464 synonyms of the 236 classes. Concept classes in the ontology were found to be coherently and consistently defined. The ontology included more than 90% of the concepts that appeared in social media posts on fertility policies. Average scores for all of the criteria for structural and representations layers exceeded 4 on a 5-point scale. Violence and abuse (socioeconomic and cultural factor) and flexible working arrangement (fertility policy) were weak signals, suggesting that they could increase rapidly in the future.

Conclusions: The determinants-of-fertility ontology developed in this study can be used as a framework for collecting and analyzing social media data on fertility issues and detecting future signals of fertility issues. The future signals identified in this study will be useful for policy makers who are developing policy responses to low fertility.

(J Med Internet Res 2021;23(6):e25028) doi:10.2196/25028

KEYWORDS

ontology; fertility; public policy; South Korea; social media; future; infodemiology; infoveillance

Introduction

South Korea has the lowest fertility rate in the world. According to the Organization for Economic Cooperation and Development (OECD), the total fertility rate (TFR) in South Korea peaked in 1970 at 4.53 and subsequently declined to 1.30 in 2001 [1]. According to Statistics Korea [2], in 2018 the TFR fell to 0.98, below the critical level of 1.

In an attempt to increase the TFR, in 2005 the Korean government enacted the Basic Law on Low Fertility and Aging Society, and the Ministry of Health and Welfare in collaboration with other government agencies established 5-year plans. The First Basic Plan for Low Fertility and Aging Society (2006-2010) was initiated to establish a foundation for the government from which to proactively respond to the low fertility and aging population. The second and third of these plans (2011-2020) were pursued with the aim of increasing the TFR and successfully responding to the increasingly aging society [3-5]. However, despite the efforts of the government over the past 15 years, the TFR in South Korea remains the lowest in the world. In order to achieve the desired policy outcome of increasing the fertility rate, the government needs to continuously identify current issues related to fertility as well as those that will arise in the future based on the detection of future signals [6].

Governments around the world are increasingly seeking ways to detect future signals of policy implications so they can respond to the various challenges that countries face in a timely and effectively manner [6]. Government foresight programs such as the UK National Horizon Scanning Centre [7] and the Finland national foresight system [8] are monitoring future signals for policy making. Future signals are signals that are not currently mainstream but are useful for predicting changes in the future. Ansoff [9] defined such signals as weak signals, referring to small opinions or symptoms with unusual patterns of future changes. Weak signals are signs that do not impact the present but can develop into strong signals and then subsequently into a trend or megatrend in the future. Thus, the issues that will be important in the future may be predicted by detecting weak signals [10].

One approach to predicting future signals is to harness intuitive judgment by experts; however, this is both time-consuming and costly [11,12]. Since the volume of textual data and influence of public opinion on social media are increasing both rapidly and continuously, there have been attempts to detect future signals using social media data [12] in various areas, including the use of solar cells [12], school bullying [13], and health and welfare policies [14]. Policy makers can apply similar approaches to social media data to detect future signals of fertility issues in order to gain valuable insights and make better policy strategies toward increasing the fertility rate in South Korea [14].

Social media data are written in various forms and are both unstructured and noisy [15]. Analyzing such unstructured data requires not only a structured framework expressing a systematic domain classification and terminology, but also knowledge of the semantic relationships between concepts [16]. A framework based on semantic analysis is required to extract information from social media data that will be valuable to government policy makers [17]. In this study we propose an ontology as a framework for the analysis of social media data.

An ontology expresses shared concepts and their relationships in a specific domain [18], and it can be used as a framework for analyzing unstructured social media data since it systematically expresses knowledge as a set of concepts in a domain and incorporates the semantics of those concepts [19-21]. An ontology that includes terminology with synonyms of the ontology class concepts was found to be useful for analyzing the language commonly used by the general public on social media [20,21]. However, an ontology with terminology representing the determinants of fertility has yet not been developed.

This study aimed to (1) develop an ontology with terminology for collecting and analyzing social media data on the determinants of fertility, (2) determine the description logics (DL), content coverage, and structural and representational layers of the ontology, and (3) use the ontology with terminology to detect future signals of fertility issues in social data posted in Korean.

Methods

Ontology and Terminology Development

An ontology for describing the determinants of fertility, called the determinants-of-fertility ontology, was developed based on the Ontology Development 101 methodology [22] in 5 steps, as described below.

Step 1. Determining the Domain and Scope of the Ontology

The aim of the determinants-of-fertility ontology developed in this study was to analyze social media data posted by consumers, not by health care professionals. Thus, we limited the scope of the ontology to the individual, social, economic, cultural, and policy factors of fertility in the domain of the consumer. The physiological, clinical, and therapeutic factors of fertility in the domain of health care professionals were excluded. The specific domain and scope of this ontology was determined by creating competency questions (CQs) that the ontology must be able to answer [23]. Since fertility is affected by multilevel factors [24], the domain and scope of the ontology were determined based on the ecological model [25]. CQs were extracted on the reproductive decisions of women from a report on low fertility in OECD countries [26] and from a research report on the causes of low fertility in South Korea [27], such as “What are the personal factors that influence a woman’s decision to have a child?” and “What are the Korean government’s policies for overcoming low fertility?” The CQs were also subsequently used for evaluating the ontology.

Step 2. Considering Reusing Existing Ontologies

We identified existing ontologies and conceptual frameworks representing fertility by searching PubMed, Google Scholar, and BioPortal [28] using the keywords “fertility,” “childbirth,” “low fertility,” “fertility ontology,” “childbirth ontology,” and “low-fertility ontology.” This search process identified an ontology representing genes associated with infertility, but this ontology was not appropriate for this study since it only included genetic factors related to infertility.

Step 3. Extracting Important Terms in the Ontology

We extracted terms from the literature that were consistent with the domain and scope of the ontology. The literature reviewed included reports on fertility, determinants of fertility, low fertility, and policy responses to low fertility published by the OECD [29], United Nations Population Division [30], and United Nations Population Fund [31], which are international organizations dealing with fertility issues jointly in countries around the world. We also searched the literature using the keywords “childbirth,” “fertility,” “determinants of fertility,” and “low fertility rate.” The reviewed literature included research papers on individual socioeconomic factors affecting low fertility and policies in countries experiencing low fertility problems such as South Korea, Japan, Singapore, Spain, Portugal, the United States, and France. Additional terms were extracted by reviewing social media posts and national surveys on fertility issues.

Step 4. Defining the Classes and Class Hierarchy

The classes of the ontology and their hierarchy were defined using a top-down approach. The superclasses of the ontology and their relationships were constructed by integrating an ecological model [25].

Step 5. Defining the Internal Structure of Classes

The internal structure of the ontology classes was defined by adding the properties of the classes, the value of the properties, and the value type using the entity-attribute-value (EAV) model. Entities refer to the concepts covered in the determinants of fertility, attributes are characteristics of entities, and value sets comprise the set of values that an attribute can have. Attributes and values were extracted from the questionnaires of the Korean National Survey on Dynamics of Marriage and Fertility [32] and the Korean Longitudinal Survey of Women and Families [33]. The ontology also included terminology with a list of synonyms for classes, attributes, and values. We formally represented the determinants-of-fertility ontology using open-source Protégé software (version 5.5.0).

Ontology and Terminology Evaluation

The available methods for evaluating the quality of an ontology include those proposed by Brank et al [34], Obrst et al [35], and Vrandečić [36]. Brank et al [34] classified ontology evaluations into 4 categories: (1) approaches that compare the target ontology to a gold standard (gold standard), (2) approaches that use the target ontology in an application (application-based), (3) approaches that compare the target ontology with a source of data (data-driven), and (4) approaches that assess the target ontology by human (user-based). Since there is no gold standard available, the determinants-of-fertility ontology was evaluated using the remaining 3 approaches proposed by Brank et al [34] (ie, application-based, data-driven, and user-based evaluations). The ontology was revised based on the evaluation results.

Evaluating the DL of the Ontology

We tested the DL of the ontology by applying the ontology debugger Protégé plug-in. We also tested the DL using the DL-reasoner Protégé plug-in to determine whether the ontology generates the correct answers to the previously developed CQs. For example, the CQ “What are the personal factors that influence a women’s decision to have a child?” was converted to a DL query “IsIndividualOf some Determinants_of_fertility.” After entering this query into Protégé, we tested whether the answers to the CQ were correct. Since the determinant of the fertility class (domain) was related to the subclasses of individual (range) through the *hasIndividual* relationship, and subclasses (eg, sociodemographic data, reproductive health, and individual’s attitude) of the individual were related to the individual class through an *is-a* relationship, we could obtain the result of a DL query.

Evaluating the Content Coverage of the Ontology

The content coverage of the ontology was examined by comparing terms extracted from the bulletin board of the Korean Ministry of Health and Welfare with a list of classes and synonyms of the ontology. Both the general public and public servants are allowed to post their opinions or concerns on fertility issues and policies regarding low fertility on this bulletin board. In total, 1387 documents posted on the website by the general public and 63 posted by public servants were collected. Relevant terms in the documents were extracted using the Korean Natural Language Processing package in R software (version 3.2.1, R Foundation for Statistical Computing). Unique concepts were extracted based on the meaning of the terms and then mapped onto the ontology classes. The mapping results were reviewed by 3 experts in health informatics who had experience in ontology development [37]. Any new concepts that were identified were added to the ontology.

Evaluating the Structural and Representational Layers of the Ontology

The structural and representational layers of the ontology were evaluated by 3 experts in health informatics who had previous experience in ontology design and 2 experts in maternity nursing who had previous experience in ontology evaluation. The evaluation tool developed by Jung et al [38] was used; this tool was based on the criteria for the structural and representational layers of Kehagias et al [39]. The structural layer was evaluated using 7 items: size, hierarchy depth, hierarchy breadth, density, balance (equally developed), overall complexity, and connectivity between classes. The representational layer was evaluated using 10 items: the match between formal and cognitive semantics, consistency, clarity, explicitness, interpretability, accuracy, comprehensiveness, granularity, relevance, and description. Each item was scored on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). The structural layer was evaluated with the entire ontology based on the hierarchy of classes and relationships between the classes, and the representational layer was evaluated with each of 6 superclasses based on the EAV model of each class.

Applying the Ontology to Detect Future Signals

The ontology with terminology was used to detect future signals of fertility issues from social media data. Future signals were analyzed based on the text-mining–based weak-signal detection method of Yoon [12], as follows, using R software (version 3.5.2).

Step 1. Collecting Data

We collected posts on fertility issues written in Korean from the following 183 online channels between January 1, 2011, and June 30, 2015: 159 channels of online news, 17 message boards, 1 social networking service (Twitter), 4 internet blogs, and 2 online community services. “Low fertility” was used as a major search keyword, together with synonyms of “fertility rate decline,” “sharp decline in fertility rate,” “avoiding childbirth,” “no kids,” and “childless family.” Social media data were collected using the SK telecom’s big-data analytics platform [40]. The data were preprocessed by treating a single document as an analysis unit.

Step 2. Defining the Keywords

After extracting terms from each document, we identified the terms related to fertility issues such as socioeconomic and cultural factors and fertility policies. The future signals of fertility issues were detected using the keywords representing socioeconomic and cultural factors and fertility policies. The keywords that were semantically similar but expressed using different terms [12] were grouped into class concepts of the ontology using terminology linking terms to concepts. The top 17 most frequently encountered keywords were selected for future-signal analysis (Textbox 1), which excluded rarely used keywords that are likely to affect the average frequency and growth rate obtained in such analyses [13,41]. The document was then coded based on whether keyword was absent (=0) or present (=1) in order to check the document frequency (DF) for the occurrence of keywords.

Textbox 1. Selected keywords.

Socioeconomic and cultural factors:

- Population aging
- Economic problems
- Nuclearization of the family
- Changing perspectives about marriage
- Conservative values
- Violence and abuse
- Employment problems
- Gender inequality

Fertility policies:

- Financial support for childbirth
- Child-safety protection system
- Infrastructure for childcare support
- Maternity-leave system
- Policy public relations
- Financial support for employment security
- Flexible working arrangement
- Family-friendly work environment
- Smart work center

Step 3. Constructing Keyword Portfolio Maps

Future signals (also defined as weak signals) show abnormal patterns due to current oddities [10]. Future signals can be detected by constructing keyword portfolio maps using the frequency information and by applying a time-weighted approach to focus on recent abnormalities [12]. We constructed a type of keyword portfolio map called a keyword issue map

(KIM) using the growth rate of the degree of diffusion (DoD) for the DF of keywords. The KIM shows the extent to which future-signal topics are diffused. The DF represents how common the term is in the collected documents. Since terms that occur frequently within collected documents are more important, the DF is directly related to future signals and can be calculated as:

The DoD is the growth rate of the term occurrence expressed as a time-weighted coefficient and is also important for detecting future signals. The DoD represents how the diffusion of a term across different documents varies over time. Since the recent appearance of a term is more important than its past appearance, the DoD puts more weight on recent occurrences:

where DF_{ij} is the DF of keyword i during period j, NN_j is the total number of documents identified for period j, n is the number of periods, and tw is a time weight (previous studies have used $tw = 0.05$ [12,41]).

The KIM was generated by plotting the average DF on the x-axis and the average growth rate of the DoD on the y-axis. The quadrants of the plot were divided by the medians of the respective values, and so each quadrant of the KIM represented different information about present and future keywords.

Step 4. Identifying Weak-Signal Topics

Future signals were identified according to where keywords were located in the quadrants of the KIM. Keywords in the first quadrant, which represent strong signals, have a trend toward a high average DF and a high average DoD growth rate. Keywords in the second quadrant, which represent weak signals, have a low average DF but a high average DoD growth rate, and so they may increase rapidly in the future. Keywords in the third quadrant, which represent latent signals, have a low average DF and a low average DoD growth rate and are not yet significantly noticeable. Keywords in the fourth quadrant, which represent not-strong-but-well-known signals, have a high average DF but a low average DoD growth rate, and so currently exhibit a slow growth rate.

Results

Ontology and Terminology Development

A list of 10 CQs was compiled (Textbox 2) that reflect different levels of factors affecting fertility using an ecological model. The domain and scope of the ontology were determined based on CQs for the determinants of fertility related to the individual, family, workplace, childcare and educational environment, socioeconomic and cultural environment, and public policy. Childbirth before marriage is very uncommon in South Korea, and so delayed marriage is an important factor affecting childbirth decision-making [3,4,27,42,43]. Therefore, the scope of fertility determinants included factors related to marriage delay and childbirth decision-making.

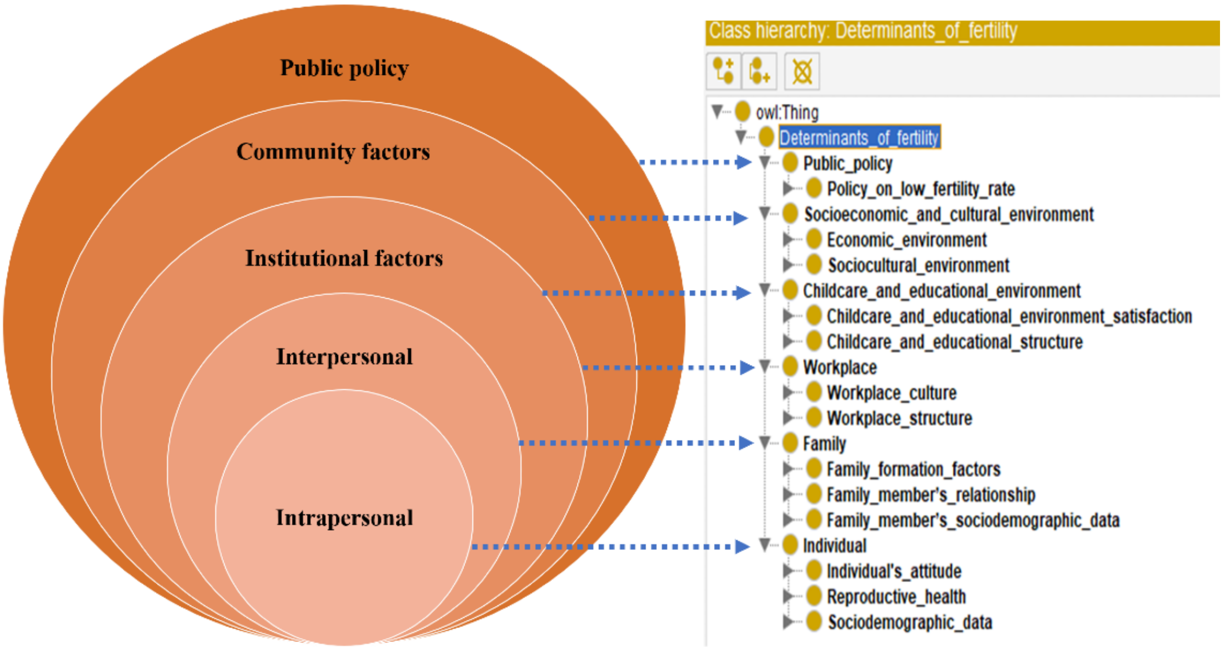
Textbox 2. Competency questions.

1. What are the personal factors that influence a woman’s decision to have a child?
2. What are the family factors that influence the decision to have a child?
3. What are the childcare factors that influence the decision to have a child?
4. What are the educational factors that influence the decision to have a child?
5. What are the workplace factors that influence the decision to have a child?
6. What are the sociocultural factors that influence the decision to have a child?
7. What are the economic factors that influence the decision to have a child?
8. What is the Korean government’s policy for overcoming low fertility?
9. What are the policy tasks for addressing low fertility in South Korea?
10. What are the policy targets for low fertility in South Korea?

In total, 1659 terms covering the domain and scope of the ontology were collected, and 236 unique class concepts were extracted from these terms. We defined hierarchical and attribute relationships of the classes based on the ecological model. The determinants of fertility were organized into the following levels: individual, family, workplace, childcare and educational environment, socioeconomic and cultural environment, and public policy. These 6 levels of the ontology were defined by adding not only the workplace, but also childcare and educational environment to institutional factors, which constitute the third level of the ecological model. Due to the increasing participation of women in the labor market, the workplace and childcare and educational environment are important factors influencing decisions about childbirth among women who are working [3,4,27,32,42,43]. Employed women in South Korea experience difficulties combining working with childbirth and childcare and are often faced with choosing between giving up their jobs to give birth and look after their child or continuing to work, thereby forsaking their desire to give birth and look after a child [32,42,43]. Therefore, we viewed the workplace and the childcare and educational environment as important social institutional factors and determinants of fertility.

Figure 1 shows the determinants-of-fertility ontology with classes up to the second level. The ontology consists of 6 superclasses: individual, family, workplace, childcare and educational environment, socioeconomic and cultural environment, and public policy. The individual superclass has the subclasses of women's sociodemographic data, which include age, education, employment, and religion; reproductive health, which includes sexual behavior, contraceptive use, and childbirth; and the individual's attitude, which includes their attitude toward children and marriage. The family superclass has the subclasses of the family's sociodemographic data, which include family size, family income, family expenditure, sex and age of children, number of children, spouse's age, and spouse's income; the family member's relationship, which includes gender equality, couple's intimacy, and family's life satisfaction; and family-formation factors, which include marriage cost. The workplace superclass has the subclasses of workplace structure, which includes workplace type, working hours, and employment insurance, and workplace culture, which includes parental leave availability. The childcare and educational environment superclass has the subclasses of the childcare and educational structure, which includes type, cost, service, and human resources, and childcare and educational environment satisfaction, which includes the belief of the quality. The socioeconomic and cultural environment superclass has subclasses of the sociocultural environment, which includes social value, mass media, and social change, and the economic environment, which includes economic growth. Finally, the public policy superclass has the subclass of policy on low fertility, which includes the area, legal basis, and tasks. These classes had 3 or 4 levels of hierarchy with 230 classes and 41 relationships.

Figure 1. The determinants-of-fertility ontology based on an ecological model.



We developed EAV models for the 139 lowest level class concepts. For example, *contraceptive use* had attributes of *hasType* and *hasLengthOfUse*, where the values of the *hasType* attribute were *condoms*, *contraceptive injection*, *contraceptive patch*, *pill*, *intrauterine device*, *diaphragm*, and *sterilization* and *hasLengthOfUse* had the value of *number of months*. We also developed a terminology linking synonyms for classes, attributes, and values: 90 synonyms for 236 classes, 9 synonyms for 54 attributes, and 501 synonyms for 772 values.

Ontology and Terminology Evaluation

Evaluating the DL of the Ontology

The Protégé ontology debugger program revealed that concept classes in the ontology were coherently and consistently defined. The DL reasoner showed that the ontology correctly answered all 10 CQs.

Evaluating the Content Coverage of the Ontology

The content coverage of the ontology is presented in Table 1. In total, 751 terms were extracted from the narratives posted by the general public and public servants. We extracted 532 unique concepts from the terms, of which 494 (92.9%) were included in the ontology. Examples of new concepts are test tube (in vitro fertilization), health insurance, and administrative division, and we added such new concepts to the ontology. The ontology was revised by adding 18 synonyms for classes, 17 value concepts, and 3 synonyms for values. The finalized version of the determinants-of-fertility ontology comprised 6 superclasses, 108 synonyms for 230 classes, 9 synonyms for 54 attributes, and 504 synonyms for 789 values.

Table 1. Results for the content coverage of the ontology.

Category	General public, n (%)	Public servants, n (%)	Total, n (%)
Existing concepts	416 (92.0)	93 (97.9)	494 (92.9)
New concepts	36 (8.0)	2 (2.1)	38 (7.1)
Total	452 (100)	95 (100)	532 (100)

Evaluating the Structural and Representational Layers of the Ontology

Average scores for all of the criteria for structural and representations layers exceeded 4 on a 5-point scale. The experts rated the hierarchy breadth, density, overall complexity, and connectivity criteria as strongly agree (score 5). The criterion with the lowest score was accuracy of the representation layers, with a score of 4.33 (Table 2).

Table 2. Results for the structural and representational layers of the ontology.

Criteria	Average score (range)
Structural layer	
Size	4.80 (4-5)
Hierarchy depth	4.60 (4-5)
Hierarchy breadth	5.00 (5-5)
Density	5.00 (5-5)
Balance	4.60 (4-5)
Overall complexity	5.00 (5-5)
Connectivity	5.00 (5-5)
Representational layer	
Match between formal and cognitive semantics	4.73 (4-5)
Consistency	4.50 (4-5)
Clarity	4.87 (4-5)
Explicitness	4.60 (3-5)
Interpretability	4.67 (4-5)
Accuracy	4.33 (4-5)
Comprehensiveness	4.77 (4-5)
Granularity	4.47 (3-5)
Relevance	4.83 (4-5)
Description	4.83 (4-5)

Applying the Ontology to Detect Future Signals

Table 3 lists the results for the computed DoD for each keyword for the socioeconomic and cultural factors and fertility policies. *Violence and abuse* (socioeconomic and cultural factor) and *flexible working arrangement* (fertility policy) had a low DF and high DoD growth rates. *Economic problems*, *population aging*, and *nuclearization of the family* (socioeconomic and cultural factors) and *child-safety protection system* (fertility policy) had a high DF and high DoD growth rates.

Table 3. Degree of diffusion (DoD), average DoD growth rate, and average document frequency for fertility issues.

Category and keyword	DoD[a]					Average DoD growth rate	Average DF[b]
	2011	2012	2013	2014	2015		
Socioeconomic and cultural factors							
Population aging	7463	8912	8002	4499	4503	0.088	6676
Economic problems	1637	2054	2523	1471	1503	0.214	1838
Nuclearization of the family	1178	1288	1229	667	628	0.054	998
Changing perspectives about marriage	1046	1528	1116	596	484	0.034	954
Conservative values	1150	1195	1139	576	565	0.036	925
Violence and abuse	685	800	726	461	527	0.158	640
Employment problems	515	510	436	306	208	–0.008	395
Gender inequality	286	298	319	190	133	0.027	245
Fertility policies							
Financial support for childbirth	3061	3145	2548	1573	1250	–0.015	2315
Child-safety protection system	1757	1632	1974	1241	1292	0.156	1579
Infrastructure for childcare support	1853	2209	1310	829	579	–0.061	1356
Maternity-leave system	1067	995	798	828	383	0.044	814
Policy public relations	878	883	894	648	361	0.015	733
Financial support for employment security	392	341	345	233	196	0.045	301
Flexible working arrangement	330	264	354	287	180	0.120	283
Family-friendly work environment	161	130	77	49	35	–0.146	90
Smart work center	131	114	90	50	27	–0.161	82

[a] DoD: degree of diffusion.

[b] DF: document frequency.

Figure 2 shows the KIM. The weak signals are marked with red rectangle (area A) in the second quadrant of the KIM. The strong signals are marked with blue rectangle (area B) in the first quadrant of the KIM. Table 4 presents the signal classification of keywords for socioeconomic and cultural factors and fertility policies in the KIM. The keywords classified as weak signals were *violence and abuse* (socioeconomic and cultural factor) and *flexible working arrangement* and *financial support for employment security* (fertility policies). The keywords classified as strong signals were *economic problems*, *nuclearization of the family*, and *population aging* (socioeconomic and cultural factors), and *child-safety protection system* and *maternity-leave system* (fertility policies). The keywords classified as latent signals were *gender inequality*, *employment problems*, and *changing perspectives about marriage* (socioeconomic and cultural factors), and *policy public relations*, *family-friendly work environment*, and *smart work center* (fertility policies). The keywords classified as not-strong-but-well-known signals were *conservative values* (socioeconomic and cultural factor) and *financial support for childbirth* and *infrastructure for childcare support* (fertility policies).

Figure 2. Future signal classification using the keyword issue map of fertility issues. Red rectangle (area A) indicates weak signals and blue rectangle (area B) indicates strong signals.

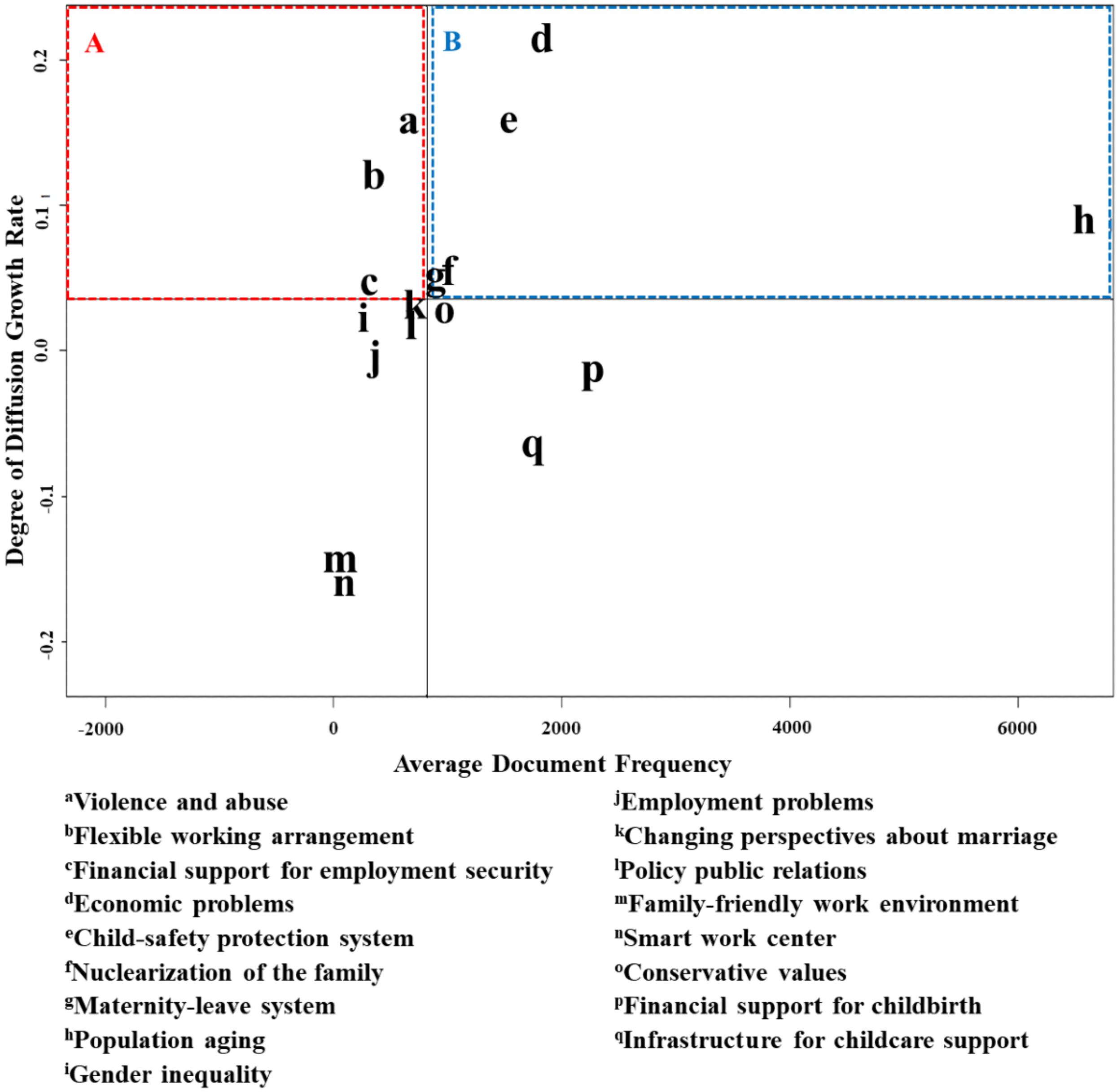


Table 4. Future signal classification of fertility-issues keywords.

Category and weak signals	Strong signals	Latent signals	Not-strong-but-well-known signals
Socioeconomic and cultural factors			
Violence and abuse	Economic problems	Gender inequality	Conservative values
—[a]	Nuclearization of the family	Employment problems	—
—	Population aging	Changing perspectives about marriage	—
Fertility policies			
Flexible working arrangement	Child-safety protection system	Policy public relations	Financial support for childbirth
Financial support for employment security	Maternity-leave system	Family-friendly work environment	Infrastructure for childcare support
—	—	Smart work center	—

[a] Not applicable.

Discussion

Principal Results

We have developed a determinants-of-fertility ontology as a framework for collecting and analyzing social media data. The ontology was evaluated in terms of the DL, content coverage, and structural and representational layers. We applied the ontology with terminology to detect future signals of fertility issues from social media data.

The developed determinants-of-fertility ontology has 6 main characteristics. First, it is the first ontology to describe the multilevel factors that affect fertility. Various factors and the complex interactions between them determine fertility [24,26,27], and expressing the fertility determinants requires an integrated view of these factors. We therefore developed the ontology by classifying the various environmental factors related to the individuals and in terms of family, childcare, workplace, and community levels from an ecological perspective.

Second, this ontology contains factors related to fertility issues that are unique to South Korea. In most cases, childbirth does not occur until after marriage in South Korea, and delaying marriage is an important factor affecting the decision to have a child [27,42,43]. Therefore, the scope of the ontology spanned individual and social factors that influence not only delayed marriage but also the decision to have children after marriage. For example, the ontology includes concepts of attitude toward marriage in the individual superclass, matters related to family formation in the family superclass, and support for starting a family in the public policy superclass.

Third, the developed ontology includes terminology with synonyms for classes such as consumer terms and abbreviations, which makes it suitable for analyzing social media data. For example, regarding *financial support for childbirth*, various terms such as *childbirth celebration money*, *childbirth grant*, *subsidy*, *baby bonus*, and *childbirth incentive* are used on social media postings. Since the developed ontology includes these terms, it can be used to collect and analyze consumer terms in social media data.

Fourth, each class of this ontology was modeled using the EAV model and included the attributes of each class and the values of those attributes. Like previous research [20,21], value sets of attributes representing the level and status of an entity included terms describing the keyword in detail in social media data. This novel characteristic of our ontology renders it capable of advanced keyword extraction and suitable for analyzing social media data.

Fifth, we ensured quality of the ontology by using a variety of evaluation methods, including the application-based, data-driven, and user-based approaches proposed by Brank et al [34]. Application-based evaluation was performed by testing the ability of the ontology to answer the CQs that cover its domain and scope. The ontology provided correct answers to all of the CQs without any errors using DL. Data-driven evaluation was performed by testing the content coverage of the ontology whereby terms extracted from social media posts on fertility issues were compared with terms included in the ontology. Medical terms related to infertility and pregnancy that appeared on social media were not included in the ontology. As a result, medical terms on social media were added to the ontology. User-based evaluation was performed by asking the experts to rate the structural consistency, errors in class relationships, and representational quality of the ontology using the evaluation criteria of the structural and representational layers [39]. The experts rated all the evaluation criteria at between 3 and 5 on a 5-point scale.

Finally, the ontology with terminology developed in this study was used as a framework to detect future signals of fertility issues from social media data. The ontology allowed us to use social media data to identify the current trends and future changes in fertility issues related to the effective implementation of policies to increase the fertility rate. These trends were *economic problems*, *child-safety protection system*, *violence and abuse*, and *flexible working arrangement*.

Economic problems and *child-safety protection system* were strong signals of fertility issues. *Economic problems* were noted to be an important topic in a survey about public perceptions of the low fertility phenomenon [44]. In a study by the Korean Ministry of Health and Welfare involving 2000 adults that examined the perceptions of low fertility and population aging, the participants reported that the main causes of low fertility were economic problems such as the economic burden of child support and education costs (60.2%) and employment instability (23.9%) [44]. Therefore, there is a need for governments to provide continuous support measures to reduce economic difficulties in family formation, childbirth, and parenting.

A *child-safety protection system* needs to be implemented as another investment toward avoiding low fertility in future generations [45]. In order to protect the health and safety of children, the government needs to actively respond to risk factors that threaten child safety such as child abuse and school violence [3,4,45-47]. Policies on child safety have been one of the most important national issues since 2013. The Korean government that came into power in 2013 has focused on protecting vulnerable groups such as children, adolescents, and women by introducing policies to deter sexual abuse, school violence, and domestic violence [46,47]. Since the government promoted these policies, both the mainstream media and social media treated them as big issues [46,47]. However, only 13.1% of married couples reportedly felt that their children were growing up in a safe and healthy environment [48], and researchers have continued to point out that the weak child-protection system remains a problem in South Korea [45,48,49]. The government therefore needs to focus more on solving the limitations and problems associated with the *child-safety protection system*.

Violence and abuse and *flexible working arrangement* were weak signals that could develop into strong signals in the future. *Violence and abuse* were previously reported as important factors for low fertility, with experts identifying changes in the social environment as being important [45,49,50]. The proportion of children (younger than 18 years) in the total population has decreased from 19.6% in 2013 to 15% in 2019 due to the low fertility rate, while the reported number of child-abuse cases has increased from 10,943 in 2012 to 36,417 in 2018 [50]. South Korea is still a highly patriarchal and authoritarian society, and a child is viewed as a subject to be parented and disciplined rather than as one with its own rights [49,50]. Thus, it is necessary to change the perception of violence and abuse toward children as social problems rather than family problems. The government needs to continually encourage policies that not only improve the fertility rate but also improve the environment that currently threatens the safety and health of babies and children.

A *flexible working arrangement* is a policy that allows workers to balance work and life [4,51]. Both the OECD and the Korean Ministry of Health and Welfare promote this as an important policy for increasing the future fertility rate in South Korea [52]. Long working hours, long commuting hours, and the culture of socializing after work in South Korea make it difficult to maintain an optimal balance between work and family [52]. Therefore, solving the low fertility problem in this country will require the government to promote policies enabling workers to have flexible working hours.

Limitations

The determinants-of-fertility ontology developed in this study comprehensively covers fertility issues relevant to the low fertility phenomenon in South Korea and will be useful for analyzing social media data. However, it is also subject to several limitations.

First, the direct and indirect effects of employment stability, job creation, housing supply, and public education on fertility [53] were not included in the public policy category of the ontology because the Korean government applies these policies separately from the policies for low fertility. A comprehensive approach covering relevant policies is needed to effectively address low fertility. Future fertility research should include policy strategies that address issues related to employment, job creation, housing supply, and public education.

Second, the synonyms of the ontology developed in this study may not include all the terms used by the general public on social media. Many of the terms used on social media are highly transient—rapidly appearing, spreading, and then disappearing [54,55]. The terms included in the ontology therefore need to be updated continuously based on those currently used by the general public.

Third, future signals of fertility issues were detected during the second phases of the policy on low fertility. Low fertility is a demographic issue that requires a long-term approach, and the policy responses of the government should be periodically reviewed and evaluated to ensure that the policies in place at a particular time point are consistent with any changes in the population and the socioeconomic and cultural environment [56]. This situation indicates the need to analyze social media data periodically using the ontology developed in this study to establish the policy direction for addressing the low fertility phenomenon by reflecting current and future trends on fertility issues as accurately and timely as possible.

Finally, only the KIM that uses the DF of keywords was used to detect future signals. Since future signals are generally subjective [10], even a careful analysis of future signals will not always yield accurate results [57]. Yoon [12] proposed a quantitative method using 2 types of keyword portfolio maps—a KIM using the DF and a keyword emergence map using the keyword frequency—to accurately detect future signals. In contrast, Lee and Park [57] proposed using both quantitative and qualitative methods to ensure the accurate detection of future signals. Therefore, we suggest performing a study of the ontology-based detection of future signals of fertility issues in South Korea that employs a quantitative method involving 2 types of keyword portfolio maps and a qualitative method involving experts.

Conclusions

A determinants-of-fertility ontology was developed in this study that comprised 6 superclasses, 230 subclasses, and 41 relationships with terminology that comprised 1464 synonyms for the 236 classes. Class concepts of the ontology were included as an EAV model and contained synonyms of the ontology classes such as consumer terms and abbreviations. The ontology can be used to analyze social media data on fertility issues. The DL, content coverage, and structural and representational layers of the ontology were evaluated. The ontology and its terminology were used to detect future signals of fertility issues in South Korea. Our novel determinants-of-fertility ontology provides a framework for collecting and analyzing social media data toward understanding which socioeconomic and cultural factors and fertility policies should be focused on in the future. The analysis of future signals revealed that *violence and abuse* (socioeconomic and cultural factor) and *flexible working arrangement* (fertility policy) were weak signals that might increase rapidly in the future. The findings of this study will help policy makers to develop effective policies for responding to the low fertility rate in South Korea based on examinations of the present and future trends.

Acknowledgments

This work was supported by grant NRF-2015R1A2A2A01008207 from the National Research Foundation of Korea funded by the Korean government.

Conflicts of Interest

None declared.

References

1. OECD. Fertility rates (indicator). URL: https://data.oecd.org/pop/fertility-rates.htm [accessed 2019-03-22]
2. Statistics Korea. Preliminary results of birth and death statistics in 2018. Daejeon; 2019. URL: http://kostat.go.kr/portal/eng/pressReleases/1/index.board?bmode=download&bSeq=&aSeq=375520&ord=2
3. Korean Government. The First Basic Plan for Low Fertility and Aging Society 2006-2010. Sejong: Korean Government; 2006.
4. Korean Government. The Second Basic Plan for Low Fertility and Aging Society 2011-2015. Sejong: Korean Government; 2011.
5. Korean Government. The Third Basic Plan for Low Fertility and Aging Society 2016-2020. Sejong: Korean Government; 2016.
6. OECD. Strategic foresight for better policies: building effective governance in the face of uncertain futures. 2019. URL: https://www.oecd.org/strategic-foresight/ourwork/Strategic%20Foresight%20for%20Better%20Policies.pdf [accessed 2020-07-06]
7. UK Government Horizon Scanning Programme team. URL: http://www.bis.gov.uk/foresight/our-work/horizon-scanning-centre [accessed 2020-07-21]
8. Prime Minister's Office Finland. URL: https://vnk.fi/en/foresight [accessed 2021-04-01]
9. Ansoff HI. Managing strategic surprise by response to weak signals. Calif Manag Rev 1975 Dec 01;18(2):21-33. [doi: 10.2307/41164635]
10. Hiltunen E. The future sign and its three dimensions. Futures 2008 Apr;40(3):247-260. [doi: 10.1016/j.futures.2007.08.021]
11. Kwon LN, Park JH, Moon YH, Lee BR, Shin YH, Kim YK. Weak signal detecting of industry convergence using information of products and services of global listed companies: focusing on growth engine industry in South Korea. J Open Innov 2018 Mar 27;4(1):10. [doi: 10.1186/s40852-018-0083-6]
12. Yoon J. Detecting weak signals for long-term business opportunities using text mining of web news. Exp Syst Appl 2012 Nov;39(16):12543-12550. [doi: 10.1016/j.eswa.2012.04.059]
13. Kim H, Han Y, Song J, Song TM. Application of social big data to identify trends of school bullying forms in South Korea. Int J Environ Res Public Health 2019 Jul 21;16(14):2596 [FREE Full text] [doi: 10.3390/ijerph16142596] [Medline: 31330879]
14. Song TM, Song J. Future signals of health and welfare policies and issues using social big data. J Health Info Stat 2016 Nov;41(4):417-427. [doi: 10.21032/jhis.2016.41.4.417]
15. Gundecha P, Liu H. Mining social media: a brief introduction. In: Mirchandani P, Smith J, Greenburg H, editors. Tutorials in Operations Research: New Directions in Informatics, Optimization, Logistics, and Production. Hanover: INFORMS; 2012:1-17.
16. Yu EJ, Kim JC, Lee CY, Kim NG. Using ontologies for semantic text mining. J Inf Syst 2012 Sep 30;21(3):137-161. [doi: 10.5859/KAIS.2012.21.3.137]
17. Belkahla Driss O, Mellouli S, Trabelsi Z. From citizens to government policy-makers: social media data analysis. Gov Inf Q 2019 Jul;36(3):560-570. [doi: 10.1016/j.giq.2019.05.002]
18. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing? Int J Human Comput Stud 1995 Nov;43(5-6):907-928. [doi: 10.1006/ijhc.1995.1081]
19. Sam KM, Chatwin CR. Ontology-based text-mining model for social network analysis. 2012 Presented at: IEEE International Conference on Management of Innovation & Technology; 2012 Jun 11-13; Bali, Indonesia p. 226-231. [doi: 10.1109/icmit.2012.6225809]
20. Jung H, Park HA, Song TM. Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. J Med Internet Res 2017 Jul 24;19(7):e259 [FREE Full text] [doi: 10.2196/jmir.7452] [Medline: 28739560]
21. On J, Park HA, Song TM. Sentiment analysis of social media on childhood vaccination: development of an ontology. J Med Internet Res 2019 Jun 07;21(6):e13456 [FREE Full text] [doi: 10.2196/13456] [Medline: 31199290]
22. Noy N, McGuinness D. Ontology development 101: a guide to creating your first ontology. Stanford: Stanford University; 2001. URL: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf [accessed 2015-05-01]
23. Grüninger M, Fox M. Methodology for the design and evaluation of ontologies. 1995 Presented at: Workshop on Basic Ontological Issues in Knowledge Sharing, in IJCAI-95; 1995; Montreal p. 1-10 URL: http://stl.mie.utoronto.ca/publications/gruninger-ijcai95.pdf
24. Bae GI, Kim KS. A study on the influence of ecological factors on the intension of childbirth. Korean Assoc Fam Relat 2011 Jul;16(2):123-149 [FREE Full text]
25. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
26. Sleebos J. Low fertility rates in OECD countries: facts and policy responses. OECD Labour Market and Social Policy Occasional Papers, No. 15. Paris: OECD Publishing; 2003. URL: https://www.oecd-ilibrary.org/low-fertility-rates-in-oecd-countries_5lgsjhvj7qg7.pdf?itemId=%2Fcontent%2Fpaper%2F568477207883&mimeType=pdf [accessed 2015-05-02]
27. Lee SS, Shin IC, Cho NH, Kim HK, Jeong YS, Choi EY. Cause of Low Fertility and Policy Responses. Sejong: Korea Institute for Health and Social Affairs; 2005.

28. National Center for Biomedical Ontology. URL: https://bioportal.bioontology.org/ [accessed 2019-01-03]
29. OECD. URL: http://www.oecd.org/ [accessed 2019-01-13]
30. United Nations Population Division. URL: https://www.un.org/development/desa/pd/ [accessed 2019-01-18]
31. United Nations Population Fund. URL: https://www.unfpa.org/ [accessed 2019-08-01]
32. Lee SS. 2005 National Survey on Dynamics of Marriage and Fertility. Sejong: Korea Institute for Health and Social Affairs; 2005.
33. Joo JS, Song CS, Son CK, Lim CS, Kim YT. Korean Longitudinal Survey of Women and Families: The 2012 Annual Report. Seoul: Women's Development Institute; 2012.
34. Brank J, Marko G, Dunja M. A survey of ontology evaluation techniques. 2005 Presented at: Proceedings of the ACM SIKDD conference on data mining and data warehouses; 2005; Ljubljana p. 166-170 URL: https://ai.ia.agh.edu.pl/_media/pl:miw:2009:brankevaluationsikdd2005.pdf
35. Obrst L, Ceusters W, Mani I. The evaluation of ontologies. In: Baker C, Cheung K, editors. Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences. Boston: Springer; 2007:139-158.
36. Vrandečić D. Ontology evaluation. In: Staab S, Studer R, editors. Handbook on Ontologies. Heidelberg: Springer; 2009:293-313.
37. Park HA, Lundberg C, Coenen A, Konicek D. Evaluation of the content coverage of SNOMED CT representing ICNP seven-axis version 1 concepts. Methods Inf Med 2011 Jan 18;50(5):472-478. [doi: 10.3414/ME11-01-0004] [Medline: 21947166]
38. Jung H, Park HA, Song TM. Development and evaluation of an adolescents' depression ontology for analyzing social data. Stud Health Technol Inform 2016;225:442-446. [Medline: 27332239]
39. Kehagias D, Papadimitriou I, Hois J, Tzovaras D, Bateman J. A methodological approach for ontology evaluation and refinement. 2008 Jun 26 Presented at: The second International ASK-IT Final Conference; 2008; Nürnberg URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.511.6332&rep=rep1&type=pdf
40. SK telecom. URL: https://www.sktelecom.com [accessed 2021-05-28]
41. Park C, Cho S. Future sign detection in smart grids through text mining. Energy Procedia 2017 Sep;128:79-85. [doi: 10.1016/j.egypro.2017.09.018]
42. Kim YO. A dynamic study of women's labor market transitions: career interruptions and its determinants. Korea J Popul Stud 2002;25(2):5-40.
43. Lee SS, Jeong KH, Hwang NM, Yoon SM, Kim SV, Lee YK. A study on the performance indicators of the Second Basic Plan for Low Fertility and Aging Society. Sejong: Korea Institute for Health and Social Affairs; 2011.
44. Ministry of Health and Welfare. 2011 National Survey of People's Perception on Population Aging of Fertility Rates. Sejong: Ministry of Health and Welfare; 2011.
45. Ryu JH. Child maltreatment and improvement direction for child protection system. Health Welfare Policy Forum 2017;247:5-23.
46. Kim DW, Park JH. A study on the current status and policy recommendations to sexual, school, and domestic violence. J Comm Saf Sec Envir Des 2014;5(1):6-31.
47. Kang SY, Park OB. A comparative study on recognition of effective safety for 'Four Social Evils'. Korean Crim Psychol Rev 2015;11(1):25-52.
48. Lee SS, Lee JH, Kim SJ. An Evaluation on the Performance of Policy Response to Low Fertility and Aging Society. Sejong: Korea Institute for Health and Social Affairs; 2014.
49. Kim JH. Population policy outlook 2018. Health Welfare Policy Forum 2018;255:61-74.
50. Ministry of Health and Welfare, National Child Protection Agency. Child abuse and neglect Korea 2018. Sejong: Ministry of Health and Welfare; 2019.
51. Korea Ministry of Government Legislation. The Labour Standard Act Article 51 and 52. Sejong: Korea Legislation Research Institute; 2019. URL: https://elaw.klri.re.kr/kor_mobile/viewer.do?hseq=46242&type=sogan&key=6
52. OECD. Rejuvenating Korea: policies for a changing society 2019. Paris: OECD Publishing; 2019.
53. Chung SH. Policy responses to low fertility and its problems. Korea J Popul Stud 2015;38(2):113-134.
54. Baron N. Language of the internet. In: Farghali A, editor. The Stanford Handbook for Language Engineers. Standford: CSLI Publications; 2003:59-127.
55. Yoon G. A study on the method of word formation through newly-coined words collected from twitter. Korean Semant 2013;42:537-555.
56. Salas R. Reflections on Population 2nd Edition. New York: Pergamon Press; 1985.
57. Lee YJ, Park JY. Identification of future signal based on the quantitative and qualitative text mining: a case study on ethical issues in artificial intelligence. Qual Quant 2017 Oct 12;52(2):653-667. [doi: 10.1007/s11135-017-0582-8]

Abbreviations

CQ: competency question
DF: document frequency

DL: description logics
DoD: degree of diffusion
EAV: entity-attribute-value
KIM: keyword issues map
OECD: Organization for Economic Cooperation and Development
TFR: total fertility rate

Edited by R Kukafka; submitted 18.10.20; peer-reviewed by R Poluru, Z He; comments to author 24.02.21; revised version received 13.04.21; accepted 25.04.21; published 14.06.21.

Please cite as:
Lee JH, Park HA, Song TM
A Determinants-of-Fertility Ontology for Detecting Future Signals of Fertility Issues From Social Media Data: Development of an Ontology
J Med Internet Res 2021;23(6):e25028
URL: https://www.jmir.org/2021/6/e25028
doi:10.2196/25028
PMID:34125068

Original Paper

Internet Interest in Colon Cancer Following the Death of Chadwick Boseman: Infoveillance Study

Hiten Naik[1*], MD; Maximilian Desmond Dimitri Johnson[2*], BASc; Michael Roger Johnson[3*], PhD

[1]Department of Medicine, University of British Columbia, Vancouver, BC, Canada

[2]Integrated Engineering, University of British Columbia, Vancouver, BC, Canada

[3]Beedie School of Business, Simon Fraser University, Vancouver, BC, Canada

[*]all authors contributed equally

Corresponding Author:
Michael Roger Johnson, PhD
Beedie School of Business
Simon Fraser University
500 Granville Street
Vancouver, BC, V6C 1W6
Canada
Phone: 1 6045370293
Email: mjohn@sfu.ca

Abstract

Background: Compared with White Americans, Black Americans have higher colon cancer mortality rates but lower up-to-date screening rates. Chadwick Boseman was a prominent Black American actor who died of colon cancer on August 28, 2020. As announcements of celebrity diagnoses often result in increased awareness, Boseman's death may have resulted in greater interest in colon cancer on the internet, particularly among Black Americans.

Objective: This study aims to quantify the impact of Chadwick Boseman's death on web-based search interest in colon cancer and determine whether there was an increase in interest in regions of the United States with a greater proportion of Black Americans.

Methods: We conducted an infoveillance study using Google Trends (GT) and Wikipedia pageview analysis. Using an autoregressive integrated moving average algorithm, we forecasted the weekly relative search volume (RSV) for GT search topics and terms related to colon cancer that would have been expected had his death not occurred and compared it with observed RSV data. This analysis was also conducted for the number of page views on the Wikipedia page for colorectal cancer. We then delineated GT RSV data for the term *colon cancer* for states and metropolitan areas in the United States and determined how the RSV values for these regions correlated with the percentage of Black Americans in that region. Differences in these correlations before and after Boseman's death were compared to determine whether there was a shift in the racial demographics of the individuals conducting the searches.

Results: The observed RSVs for the topics *colorectal cancer* and *colon cancer screening* increased by 598% and 707%, respectively, and were on average 121% (95% CI 72%-193%) and 256% (95% CI 35%-814%) greater than expected during the first 3 months following Boseman's death. Daily Wikipedia page view volume during the 2 months following Boseman's death was on average 1979% (95% CI 1375%-2894%) greater than expected, and it was estimated that this represented 547,354 (95% CI 497,708-585,167) excess Wikipedia page views. Before Boseman's death, there were negative correlations between the percentage of Black Americans living in a state or metropolitan area and the RSV for *colon cancer* in that area (r=−0.18 and r=−0.05, respectively). However, in the 2 weeks following his death, there were positive correlations between the RSV for colon cancer and the percentage of Black Americans per state and per metropolitan area (r=0.73 and r=0.33, respectively). These changes persisted for 4 months and were all statistically significant (P<.001).

Conclusions: There was a significant increase in web-based activity related to colon cancer following Chadwick Boseman's death, particularly in areas with a higher proportion of Black Americans. This reflects a heightened public awareness that can be leveraged to further educate the public.

(J Med Internet Res 2021;23(6):e27052) doi:10.2196/27052

KEYWORDS

colon cancer; Google; Wikipedia; infodemiology

Introduction

Background

On August 28, 2020, it was announced through Chadwick Boseman's Twitter account that he had died of colon cancer at the age of 43 years. Boseman was a Black American actor who was perhaps best known for playing the Black Panther in the Marvel Cinematic Universe. The announcement was shocking, as he was a seemingly healthy young man whose cancer diagnosis was not disclosed to the public. Boseman is an icon in the Black community. In addition to his lauded role as the fictional Black Panther, he also played several prominent, historically impactful Black Americans in his movies, such as the baseball player Jackie Robinson, funk musician James Brown, and civil rights activist and supreme court justice Thurgood Marshall. The surprising news of his death was shared widely on social media, and the tweet soon became the most liked in Twitter history [1].

Boseman's death was a tragic loss, but it may have had a beneficial aspect. The degree of attention that it spawned may result in a greater awareness about colon cancer, a disease that is curable if caught early by established screening methods. Prior research has demonstrated that the announcement of cancer or related diagnosis by a celebrity results in a heightened interest in that cancer and that this can extend to greater participation in screening and primary prevention [2-7]. Probably the most well-known case is the *Angelina effect*, a term coined by Time magazine [8] to describe the impact that Angelina Jolie had in 2013 when she publicly disclosed in a New York Times editorial that she had undergone a risk-reducing bilateral mastectomy after learning that she was a BRCA1 (Breast Cancer Gene 1) mutation carrier [9]. Public interest in breast cancer has soared, and there has been a subsequent increase in genetic testing [2,7]. Similarly, it is possible that Boseman's death will lead to increased uptake of colon cancer stool tests and screening colonoscopies in the coming years.

However, it is unclear whether this spike in interest from celebrity-related news truly has a positive impact on health-related behaviors in the long term [6,10]. A 2016 analysis of the Angelina effect revealed that although the rates of breast cancer gene testing increased following her New York Times editorial, the rates of mastectomies among those tested declined [6]. This implied that the women who underwent the genetic tests after her revelation had a lower pretest probability of testing positive than those tested before. The authors concluded that celebrity announcements might not effectively target the subpopulation at the highest risk of a particular disease [6]. In contrast, the fallout of Boseman's death has the potential to be different because he was a Black American. Despite having a greater risk of colon cancer, Black Americans lag behind White Americans with respect to up-to-date screening rates [11]. If Boseman's legacy results in an increased awareness about colon cancer that leads to an increased screening in Black communities, the impact could be truly substantial in lowering the incidence of colon cancer.

Research Objectives

It is likely too early to determine whether Boseman's death has had a national effect on the rates of colon cancer screening and diagnoses. However, public interest in colon cancer can be evaluated sooner by monitoring the relative internet activity surrounding the topic. In an emerging field known as *infodemiology*, researchers have used the activity on search engines and other websites to make inferences about health issues at the population level [12-16]. This process of real-time surveillance of the supply and demand of health information has been dubbed *infoveillance* [12,13,16].

Among the most common tools for infoveillance is Google Trends (GT), a feature provided by Google that allows for quantification of Google searches for specific terms or broader topic areas [17]. For example, GT has been used to monitor outbreaks of infectious diseases [14,15,18-20], infer cancer incidences in specific geographic locations [3,21], and understand trends in health behaviors such as smoking cessation [22,23], suicide rates [24], uptake of cannabis products [25], and cancer screening [26,27]. It has also proven to be a valuable tool for tracking the impact of public health campaigns [22,28] and celebrity-related news [23,29,30]. For example, by assessing search volumes, it has been determined that there was an increased interest in condoms and HIV testing following Charlie Sheen's disclosure of HIV [29], Asperger syndrome following Greta Thurnberg appearances [31], and mental health hotlines following Demi Lovato's overdose and Anthony Bourdain's suicide [30].

Although not as widely used as GT data, quantifying the number of page views of disease-specific articles on Wikipedia has also been used to evaluate public interest in that disease. Studies have documented increases in page views in temporal relation to celebrity diagnosis and deaths. For example, a study showed that there were more page views on the Wikipedia article for strokes after Margaret Thatcher's death, on the multiple sclerosis article when Ozzy Osbourne announced his diagnosis, and on the epilepsy article when Lil Wayne was hospitalized with seizures [32]. Similarly, there was an increase in traffic on the Wikipedia article for vasculitis soon after Harold Allen Ramis died of that condition [33].

What remains under study is how this infoveillance data can be effectively leveraged to improve health outcomes for the public. In the examples cited, there is typically a spike in activity following the event in question before returning to baseline [29,31,33]. An important objective would be for policy makers to use these data to address disparities in diseases such as colon cancer, and to accomplish this, it would be valuable to understand patterns of interest for particular demographic groups such as in Black Americans. Although it is not possible to know the identities of the individuals who have elevated interest, GT allows the search volume data to be granulated down to specific geographic areas. Understanding the demographics of the areas with heightened activity following an event can help us infer which demographic groups had increased interest so that they may be targeted in public health campaigns featuring the celebrity.

Therefore, characterizing the impact of Chadwick Boseman's death on the internet would be valuable to public health leaders and colon cancer advocacy groups who aim to address racial disparities. The goals of this study are twofold. First, we use GT and Wikipedia page view data to evaluate the degree of increased interest in colon cancer and related topics and terms following Boseman's death. Second, we analyze the geographic distribution of this increased interest to determine if it disproportionately originated from Black Americans. With this information, we seek to provide a framework that could be used to develop targeted campaigns aimed at increasing screening and awareness of colon cancer, particularly among Black people in the United States.

Methods

Google Searches and Wikipedia Page Views for Colon Cancer

We used GT [34] to assess the weekly relative search volume (RSV) of topics and terms related to colon cancer in the United States. Search *terms* are specific words or phrases, whereas *topics* encompass many terms, as defined by Google, that share the same concept. The topics used were *colorectal cancer* and *colon cancer screening*. A total of 15 terms related to colon cancer were specifically chosen to represent queries related to colon cancer screening and diagnosis (*colonoscopy*, *stool test*, and *diagnosis*), symptoms and signs (*stool*, *symptoms*, *signs*, and *anemia*), risk factors (*risk*, *men*, *age*, *black (-panther)*, and *African American*), treatment (*treatment*), and prognosis (*survival*, *death*). According to Google, RSV is calculated by dividing the number of searches for a particular term or topic by the total searches based on geography and time range. The resulting numbers are then scaled from 0 to 100 to provide a relative indication of the popularity of the search query. For each topic and term, weekly RSV data points were downloaded for the United States for a period of 2 years before Boseman's death to 3 months after (week of September 2, 2018, to week of November 29, 2020). Each topic and term was used as an independent search query for GT.

We also used page views analysis [35] to extract data regarding the number of views of the English Wikipedia page for colorectal cancer between 2 years before Boseman's death and 3 months after (August 28, 2018, to November 29, 2020). The daily views of this page were then normalized using the total number of views of all English Wikipedia pages for each day.

On the basis of the historical data 2 years before Boseman's death, we conducted quasi-experimental analyses by forecasting the weekly Google RSV and the daily number of Wikipedia page views that would be expected if his death had not occurred and compared it with the observed values. The forecasts were generated with 95% bootstrapped CIs using an advanced autoregressive integrated moving average algorithm in the forecast package [36] of the R software (version 4.0.3; R Foundation).

Geographic Patterns in Searches for Colon Cancer

To investigate a possible shift in demographics relative to search volume for colon cancer, we analyzed Google RSV data further characterized by state and metropolitan regions. The *Interest in subregion* option on GT generates RSV values for each specific state for a search term during a given period. The RSV is scaled from 0-100 based on the popularity relative to the total number of Google searches performed during a specified time in the specified states where there is sufficient data. Instead of by state, this can be done separately by metropolitan regions in the United States, which on GT is delineated based on Nielsen designated market area (DMA) boundaries. We used the search term *colon cancer* and extracted RSV data for the period of 2 years before Boseman's death (August 28, 2018, to August 27, 2020) and periods up to 4 months following his death (August 28, 2020 to December 28, 2020).

We then compared these RSV values with the percentages of Black, White, and Hispanic and Latino American populations for each of the states and metropolitan areas and conducted correlation analyses. These correlations were expressed as the Pearson coefficient of correlation (r), and correlations were compared between the period 2 years before Boseman's death and time intervals up to 4 months following his death. This comparison was performed using a statistical test to compare 2 overlapping correlations based on dependent groups [37]. Visualizations (scatter plots and maps) were created using Tableau Desktop (Tableau Software LLC). For mapping purposes, each state or metropolitan region was shaded based on how the RSV values differed from the mean RSV value of all state or metropolitan regions during that period to highlight relative increases and decreases before and after Boseman's death.

Demographic information regarding the percentage of each race in each state and DMA was derived from the 2019 US Census data. State-level data were extracted from the United States Census Bureau website [38]. Data regarding the percentage of each race in each DMA were purchased from Cubit Planning Inc (Austin) through a custom data request. This company had previously developed a strategy to derive this information by extracting county-specific demographic data from US Census data sets and then determining which county or counties comprised each DMA [39].

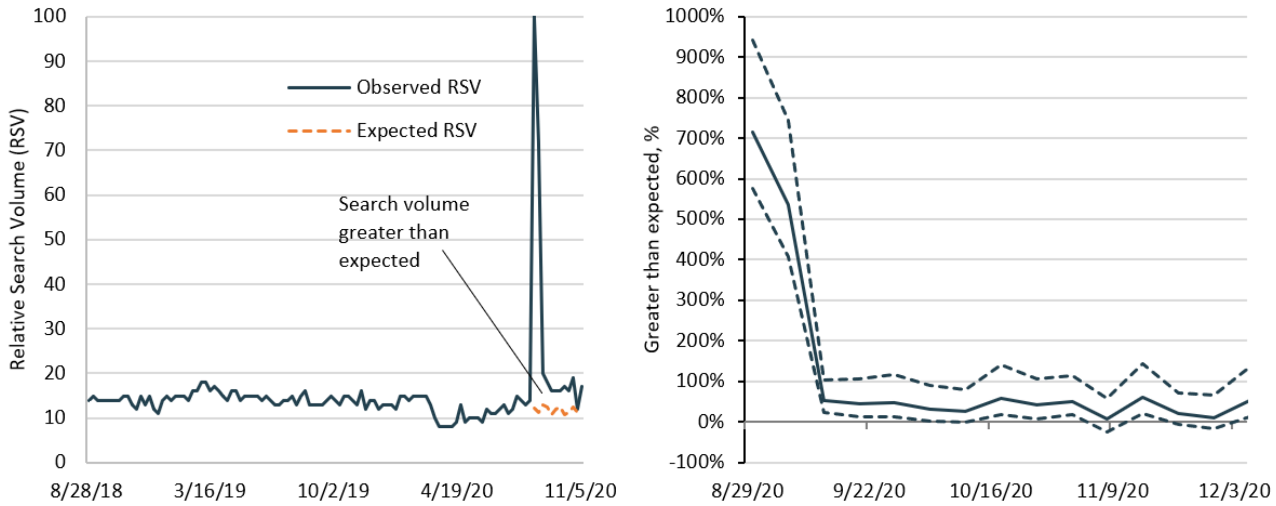
Results

Google Searches for Colon Cancer

Following August 28, 2020, GT observed RSV values for the search term *colon cancer* and the search topic *colon cancer screening* increased beyond the expected RSV by 716% and 707%, respectively (Figure 1). In the case of the search term *colon cancer*, observed RSV values remained significantly greater than expected for at least 2 months, during which searches were on average 189% (95% CI 132%-291%) greater than expected. By the third month, the average observed RSV was 118% (95% CI 72%-201%) greater than expected. Searches for the topic *colon cancer screening* were also significant 2 months following August 28, 2020, with an average of 261% (95% CI 54%-1252%) greater than expected (Figure 1). Similarly, by the third month, the average observed RSV remained significantly greater than expected, with an average of 256% (95% CI 35%-814%).

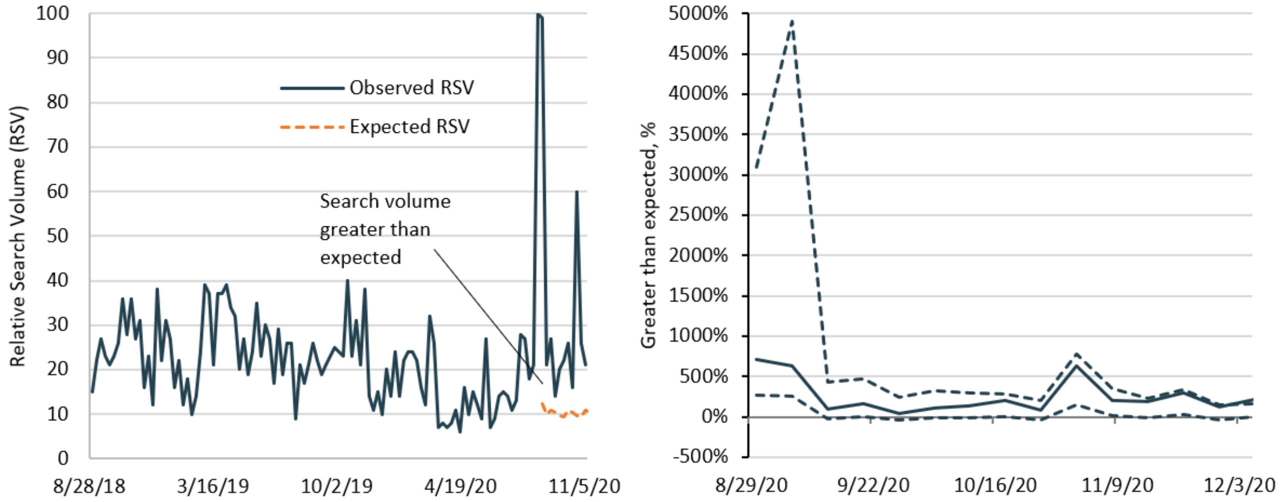
When analyzing related GT search terms for colon cancer, 10 of 15 had mean observed RSVs significantly higher than expected within the first 2 months (Figure 2), and of these, 3 (*colon cancer signs*, *colon cancer survival*, and *colon cancer symptoms*) remained significantly greater than expected within the third month.

Figure 1. Greater than expected relative search volumes for colon cancer following Chadwick Boseman’s death. RSV: relative search volume.



A) Observed search volume for “colon cancer”

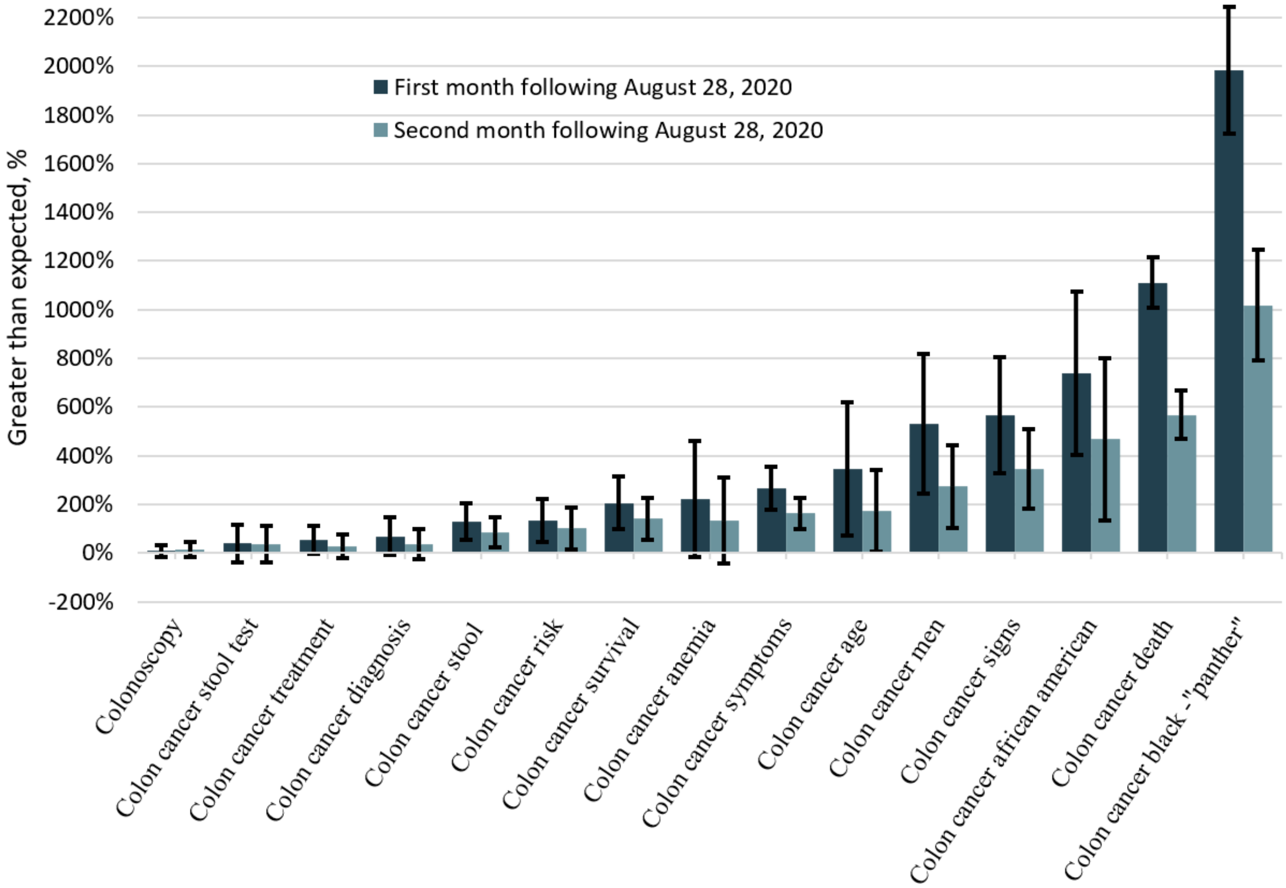
B) Excess search volume for “colon cancer”



C) Observed search volume for “colon cancer screening”

D) Excess search volume for “colon cancer screening”

Figure 2. Greater than expected relative search volumes for Google search terms related to colon cancer following Chadwick Boseman's death.

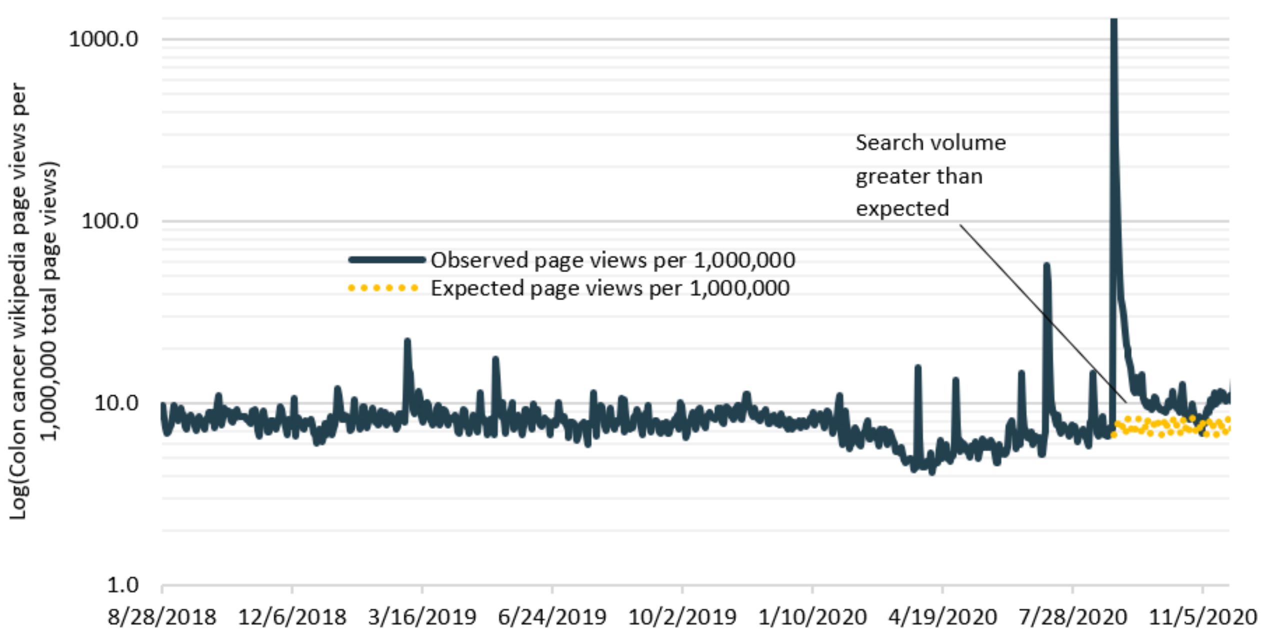


Wikipedia Page Views for Colon Cancer

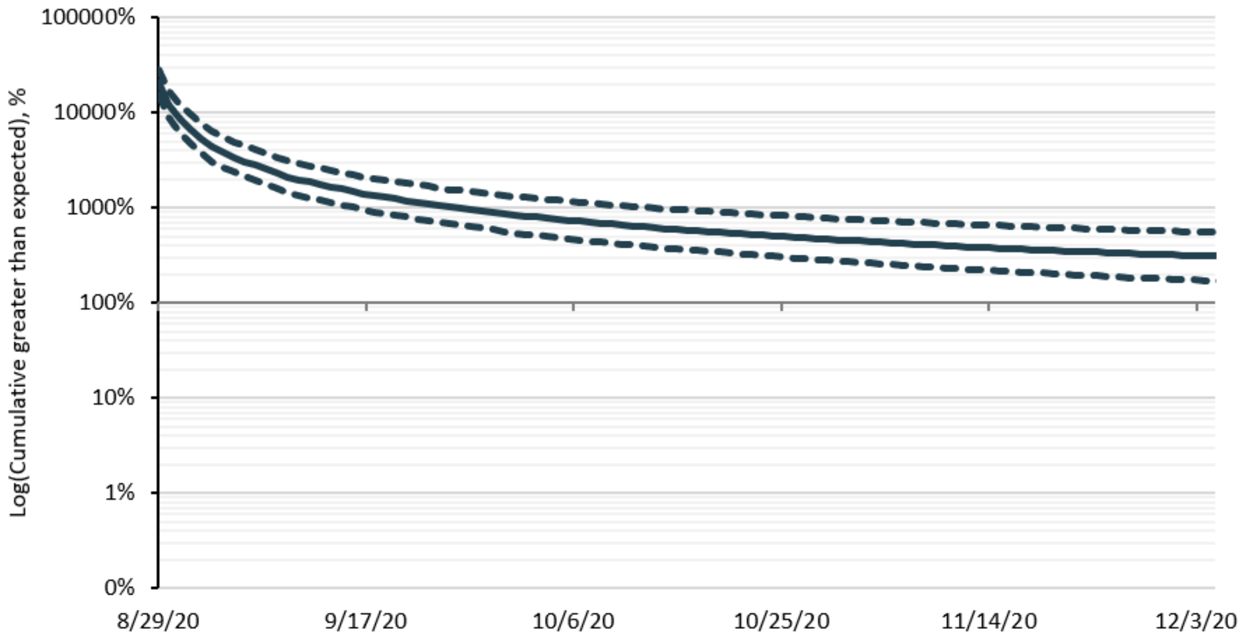
Daily page views for the English Wikipedia page *colorectal cancer* peaked at 1386.7 per million total views or 20,649% (95% CI 15,188%-27,901%) greater than expected on August 29 before declining (Figure 3). As shown in Figure 3, the cumulative excess search volume during this period was on average 1979% (95% CI 1375%-2894%) greater than expected. Altogether, it was estimated that there were 547,354 (95% CI 49,7708-58,5167) excess Wikipedia page views during the 2 months following the news of Boseman's death.

Figure 3. Excess page views for the Wikipedia page colorectal cancer following Chadwick Boseman’s death.

A) Colon cancer Wikipedia page views per million total Wikipedia page views



B) Cumulative excess search volume % for colon cancer Wikipedia page views 3 months post death



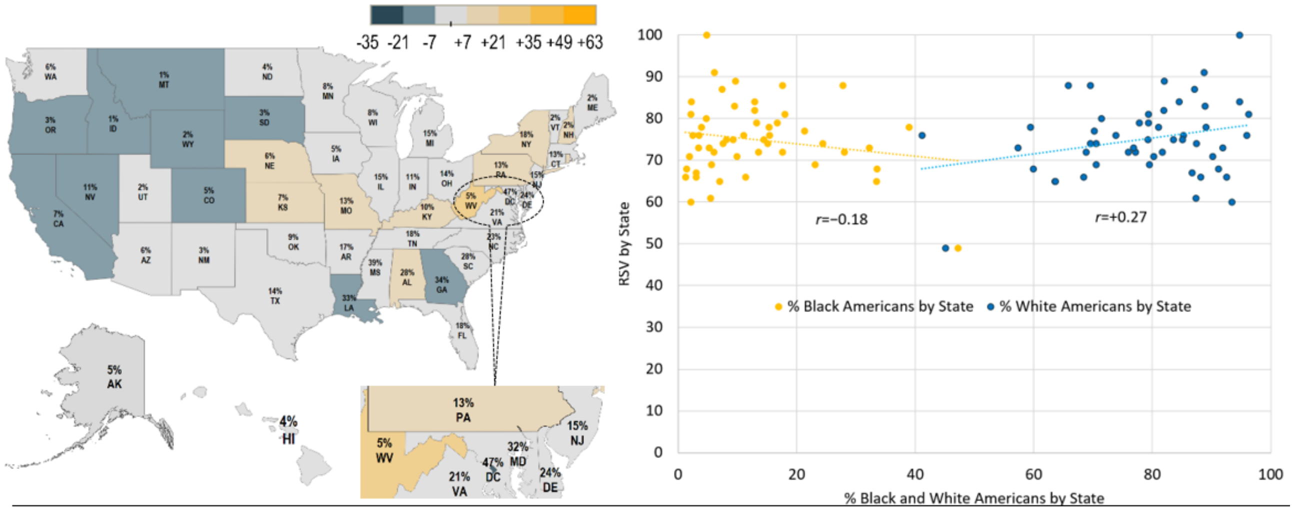
Geographic Patterns in Searches for Colon Cancer

As shown in Figure 4, for 2 years before August 28, 2020, RSV for the search term *colon cancer* was greatest among states in the northeast extending through the Midwest to Nebraska (with Alabama being the only southern state above the average RSV). In the 2 weeks following August 28, there was a geographic shift, such that states in the south tended to have more search interest. The lists of the states that had the most search interest and change in search interest following Boseman’s death are outlined in Multimedia Appendix 1, Table S1. The percentage of the Black American population in each state as per the 2019 US Census is indicated in Figure 4, and for each DMA in Multimedia Appendix 1, Figure S1.

Figure 4. Shifts in demographics and Google relative search volumes for Colon Cancer by state before and after Chadwick Boseman's death. The percentage of the Black American population by state is indicated in a smaller text on the map based on the 2019 US Census. The shading scale is based on the difference between the relative search volume value for each state and the mean relative search volume for all states during the specified period.

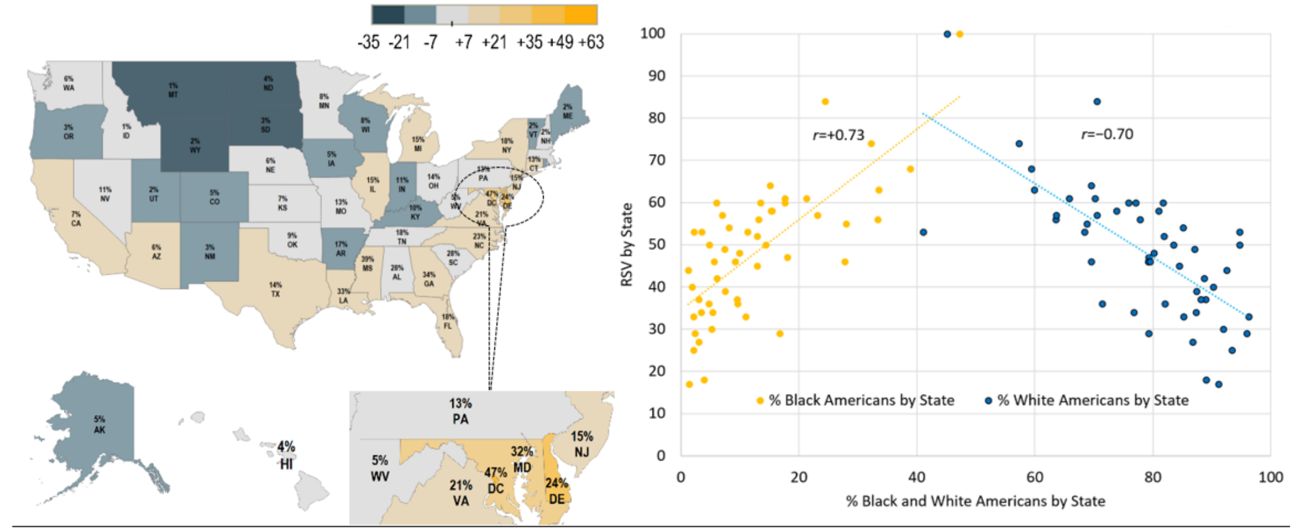
A) Before: August 2, 2018 – August 22, 2020

B) Before: August 2, 2018 – August 22, 2020



C) 2 weeks after: August 28, 2020 – November 9, 2020

D) 2 weeks after: August 28, 2020 – November 9, 2020

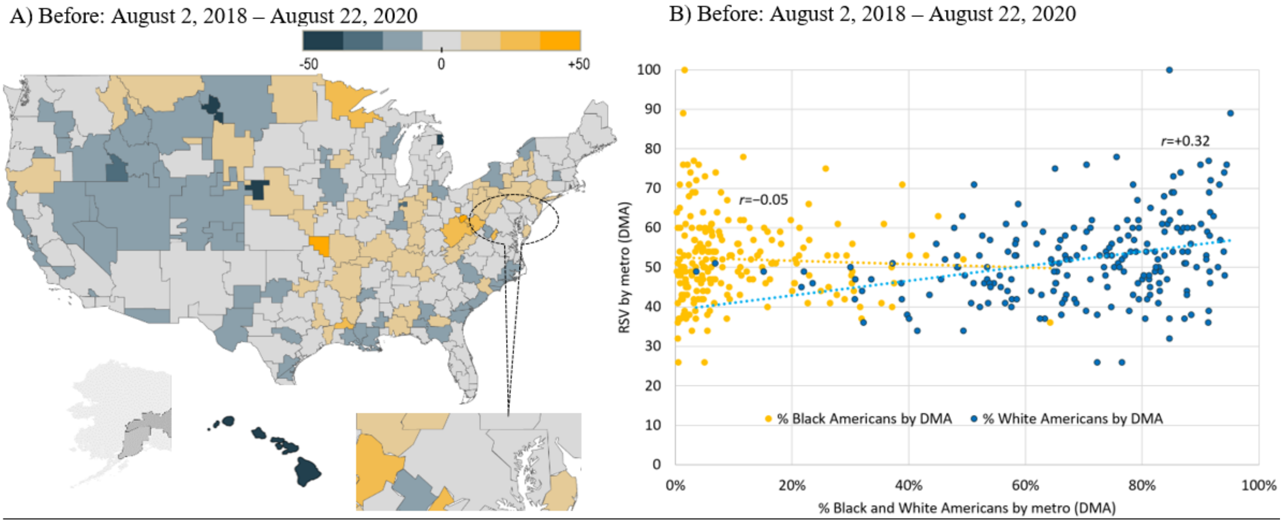


A similar geographic shift was observed when examining metropolitan areas defined using the DMA boundaries (Figure 5). To determine how this shift corresponded to changes in interest among racial groups, we conducted a series of correlation analyses. As shown in Figures 4 and 5, before August 28, 2020, there was a slight negative correlation between the RSV for colon cancer and the percentage of Black Americans per state and per metropolitan area (*r*=−0.18 and *r*=−0.05, respectively) and a slight positive relationship for the percentage of White Americans per state and per metropolitan area (*r*=0.27 and *r*=0.32, respectively). In the 2 weeks following August 28, 2020, there was a significant shift in correlation within racial demographics and RSV for *colon cancer* by state and metropolitan area, such that there was a positive correlation between the RSV for colon cancer and the percentage of Black Americans per state (*r*=0.73 and *r*=0.33, respectively) and a negative relationship for the percentage of White Americans per state and metropolitan area (*r*=−0.70 and *r*=−0.39, respectively). These changes in correlations were all highly significant (*P*<.001), relative to the corresponding correlations during the 2 years before August 28, 2020.

Figure 5. Shifts in demographics and Google relative search volumes for colon cancer by metropolitan area before and after Chadwick Boseman's death. Metropolitan areas are delineated based on Nielsen designated market areas. The shading scale is based on the difference between the relative search volume value for each metropolitan area and the mean relative search volume for all states during the specified period. DMA: designated market area.

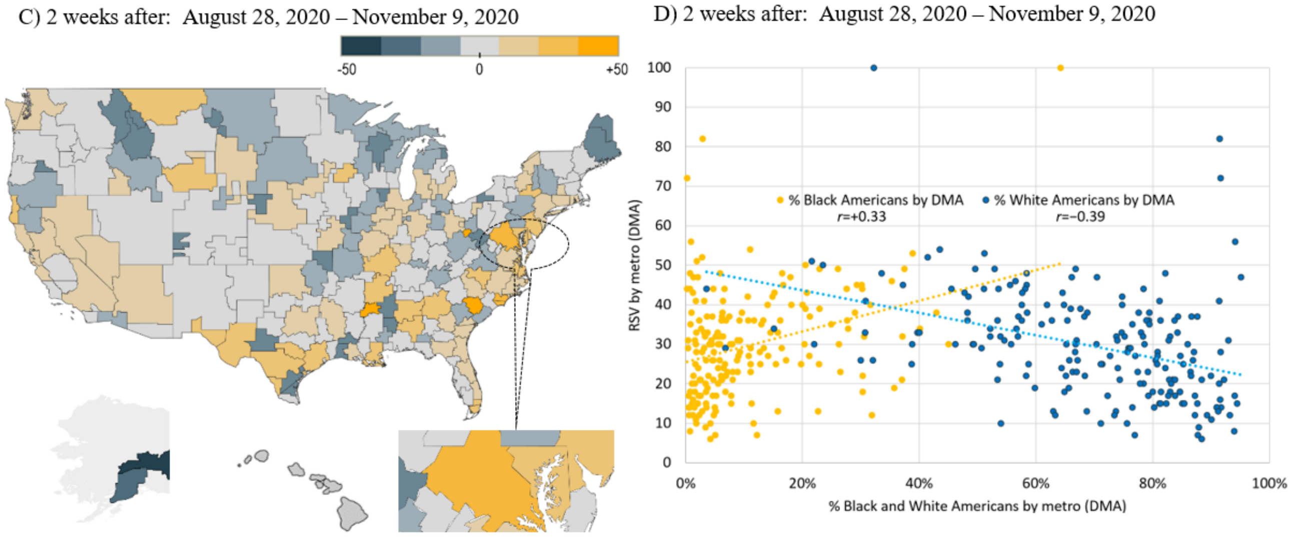
A) Before: August 2, 2018 – August 22, 2020

B) Before: August 2, 2018 – August 22, 2020



C) 2 weeks after: August 28, 2020 – November 9, 2020

D) 2 weeks after: August 28, 2020 – November 9, 2020



This analysis was repeated for periods extending up to 4 months following Boseman's death. As shown in Table 1, after August 28, 2020, the correlations remain positive between the RSV for *colon cancer* and the percentage of Black Americans by state and metropolitan area. Similarly, the percentage of White Americans and the RSV for colon cancer is negative and significantly lower after August 28, 2020, for all periods compared by state and metropolitan area. As shown in Table 1, Hispanic and Latino Americans have a similar pattern as Black Americans concerning RSV for colon cancer with time.

Table 1. Correlations between search interest in colon cancer and racial demographics.

Search terms and racial demographics	Google search dates[a]						
	Before (September 2, 2018-August 22, 2020)	1 week after (August 28, 2020-September 4, 2020)	2 weeks after (August 28, 2020-September 11, 2020)	4 weeks after (August 28, 2020-September 25, 2020)	8 weeks after (August 28, 2020-October 23, 2020)	12 weeks after (August 28, 2020-November 20, 2020)	4 months after (August 28, 2020-December 31, 2020)
***Colon cancer* by state**							
Black Americans							
Correlation coefficient, *r*	−0.18	0.51	0.73	0.68	0.62	0.69	0.65
P value	—[b]	<.001	<.001	<.001	<.001	<.001	<.001
White Americans							
Correlation coefficient, *r*	0.27	−0.69	−0.70	−0.68	−0.69	−0.70	−0.46
P value	—	<.001	<.001	<.001	<.001	<.001	<.001
Hispanic and Latino Americans							
Correlation coefficient, *r*	−0.27	0.21	0.18	0.20	0.20	0.12	0.11
P value	—	.03	.04	.03	.03	—	—
***Colon cancer* by metropolitan area (DMA[c])**							
Black Americans							
Correlation coefficient, *r*	−0.05	0.24	0.32	0.20	0.28	0.20	0.25
P value	—	.004	<.001	.002	<.001	<.001	<.001
White Americans							
Correlation coefficient, *r*	0.32	−0.27	−0.38	−0.35	−0.34	−0.31	−0.34
P value	—	<.001	<.001	<.001	<.001	<.001	<.001
Hispanic and Latino Americans							
Correlation coefficient, *r*	−0.29	0.12	0.22	0.23	0.16	0.18	0.18
P value	—	<.001	<.001	<.001	<.001	<.001	<.001

[a] Each value represents the Pearson correlation coefficient between the relative search volume and the percentage of the population for each race in the corresponding jurisdiction.

[b] N/A: not applicable.

[c] DMA: designated market area (Google Trends defines metro areas according to 210 DMAs).

Discussion

Principal Findings and Implications

In this infoveillance study, we examined the impact of Chadwick Boseman's death from colon cancer on the interest in this disease on the internet. We found that Boseman's death resulted in a dramatic and significant increase in Google searches related to colon cancer as well as in the number of page views of the colon cancer Wikipedia article. We further investigated how the increased Google search interest in colon cancer was distributed geographically by state and metropolitan areas. Our findings demonstrated a significant shift in correlation within the Black American population and RSV for colon cancer in state and metropolitan regions. This correlation remained for at least 4 months afterward.

Research studies thus far have shown that searches for colonoscopy have increased with time [27] and correlate with actual screening rates in the United States [26]. However, although current campaigns such as the colon cancer awareness month result in increased interest on the web, these campaigns do not seem to generate enough impact to result in increased screening [40]. Involving the stories of celebrities like Chadwick Boseman in these campaigns may be a more effective strategy. Previously, the disclosure of cancer diagnoses from other celebrities has resulted in higher screening rates. For example, when Kylie Minogue announced her breast cancer diagnosis in 2005, there was an uptick in mammography appointments in Australia [5]. Similarly, there was a temporary increase in pap smear rates in the United Kingdom after Jade Goody was diagnosed with cervical cancer in 2008 [4]. It is possible that Boseman's death will have a comparable, if not greater, impact on colon cancer screening rates.

Although analyzing screening and corresponding detection rates of cancers takes time, infodemiology research can provide early insights that public health officials can act upon. Multiple studies have examined the changes in interest on the web concerning a celebrity event, and several have used a quasi-experimental approach like ours, in which increased search volumes following an event are quantified in relation to counterfactual forecasted volumes if the event had not occurred [23,30,33]. For example, studies have reported that search queries regarding smoking cessation interest in Brazil increased by up to 153% in the first week after president Lula da Silva's diagnosis of laryngeal cancer [23], HIV search queries in the United States increased by 417% in the week after Charlie Sheen's disclosure [29], and there was a 257% increase in vasculitis-related queries in the United States after Harold Allen Ramis' death [33]. In contrast, our study determined that the magnitude of increased search volume attributed to the *Chadwick Boseman effect* was found to be higher, with search queries related to colon cancer screening increasing by up to 707%.

This greater volume of web-based interest in colon cancer could be attributed to Boseman's widespread popularity and the sheer surprise of his death. However, it is likely that social media played an important role as well, as unlike most other celebrity examples, Boseman's diagnosis was shared via Twitter [1]. This original tweet quickly went viral, and the news spread on multiple platforms. Our study showed that this increased social media activity translated to a greater appetite for knowledge of colon cancer itself, highlighting the value of platforms such as Twitter for public health campaigns. Several studies have demonstrated that social media–based campaigns can be successful in educating the public and increasing participation in cancer screening [41]. However, most of the campaigns studied thus far have focused on breast and prostate cancer and have been less effective at reaching racial minorities [41]. The Chadwick Boseman effect phenomenon demonstrated that the involvement of a highly followed celebrity might represent an effective strategy to expand the reach of these campaigns on the internet. Furthermore, if the celebrity represents a racial minority group like Boseman, it may facilitate the engagement of these particular groups.

Our analysis showed that there was proportionately more search volume for colon cancer in states and metropolitan areas that had proportionately more Black Americans. This reinforces how popular Boseman was in the Black American community and how his legacy could be used to help spread awareness of colon cancer among his fans. This geographic component of GT infoveillance has been leveraged in studies examining disease incidence [3,19]; however, its role in research examining public health behavior and disparities is less established. Our study adds to the infoveillance field by demonstrating that internet search trend data can not only be used to identify an increased public interest in a health-related topic but can also determine *where* that interest comes from and perhaps even from *whom*.

As we emerge from the COVID-19 pandemic, an important public health goal will be to re-engage the public in cancer screening. Research has revealed that cancer screening rates are lower and that cancers are being diagnosed at more advanced stages than that observed prepandemic [42]. In the United States, Black Americans have long had lower up-to-date screening rates for colon cancer [11], and this disparity may have become more pronounced during the pandemic. Therefore, it will be even more important to appeal to Black Americans in campaigns focused on colon cancer education and the importance of screening.

Although there was a substantial increase in internet activity regarding colon cancer following Boseman's death, the activity ultimately declined, which may have been a missed opportunity to educate the public. However, by invoking Boseman's legacy and involving living Black American celebrities in colon cancer campaigns, it may be possible to regenerate this spike in interest among Americans. This can be done on the internet but through other means, such as television commercials and the distribution of stool testing kits at the grassroots level. The GT analyses presented in our study provide a geographic framework for the jurisdictions to target in these campaigns. For example, if a public health official in the US were to start a television campaign invoking Boseman's legacy, they may wish to target regions such as Georgia and Washington DC, where there was a higher concentration of Black Americans as well as the greatest interest in colon cancer following his death.

Limitations

There are limitations of our study that are important to highlight. Internet activity for a particular condition is ultimately just a surrogate marker for the public interest. It is unknown how many different people performed the searches and who they were. Our geographic analyses suggested an increased interest in colon cancer in regions of the United States with more Black Americans. This does not provide direct evidence that more Black Americans were looking up information about colon cancer, and there may have been factors other than race that drove interest in these regions.

It is also unknown whether individuals who searched for information about colon cancer gleaned any valuable knowledge from their internet visits that would influence their future behavior. Although the Wikipedia page for colorectal cancer provides a variety of information, we were unable to capture how much of the material was actually reviewed by visitors. Furthermore, we also acknowledge that the search terms investigated may not represent all of the most common search strings used by the public.

Conclusions

This infoveillance study revealed a substantial increase in Google searches and Wikipedia page views related to colon cancer after the Black American actor Chadwick Boseman died of the disease in August 2020. A geographic analysis of states and metropolitan areas in the United States demonstrated that the surge in interest might have disproportionately come from Black Americans, a group that has a higher risk of colon cancer but lower screening rates than White Americans. Further research will be required to determine whether the temporary surge in interest following Boseman's death will translate to sustained awareness and screening rates for colon cancer.

Health care providers and public health officials can capitalize on these findings by developing awareness campaigns incorporating Boseman's legacy and involving other Black American celebrities. Although he quietly suffered from cancer himself, Boseman generously donated to multiple charities, including those supporting underprivileged Black Americans and youth affected by cancer. In his life, Boseman was known for portraying fictional and real-life Black heroes on the movie screen, but his status as a hero himself may be further cemented in the years to come.

Conflicts of Interest

None declared.

Multimedia Appendix 1
States with the highest internet search interest in "colon cancer" and the percentage of Black Americans by metropolitan area.
[DOCX File , 347 KB - jmir_v23i6e27052_app1.docx]

References

1. Pulver A. Final tweet from Chadwick Boseman's account is most liked ever on Twitter. The Guardian. 2020. URL: https://www.theguardian.com/film/2020/aug/31/final-tweet-from-chadwick-boseman-account-most-liked-ever-on-twitter [accessed 2020-12-07]
2. Liede A, Cai M, Crouter TF, Niepel D, Callaghan F, Evans DG. Risk-reducing mastectomy rates in the US: a closer examination of the Angelina Jolie effect. Breast Cancer Res Treat 2018 Sep;171(2):435-442 [FREE Full text] [doi: 10.1007/s10549-018-4824-9] [Medline: 29808287]
3. Phillips CA, Barz LA, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship between state-level Google online search volume and cancer incidence in the United States: retrospective study. J Med Internet Res 2018 Jan 08;20(1):e6 [FREE Full text] [doi: 10.2196/jmir.8870] [Medline: 29311051]
4. MacArthur GJ, Wright M, Beer H, Paranjothy S. Impact of media reporting of cervical cancer in a UK celebrity on a population-based cervical screening programme. J Med Screen 2011 Dec 07;18(4):204-209. [doi: 10.1258/jms.2011.011092] [Medline: 22156146]
5. Chapman S, McLeod K, Wakefield M, Holding S. Impact of news of celebrity illness on breast cancer screening: Kylie Minogue's breast cancer diagnosis. Med J Aust 2005 Sep 05;183(5):247-250. [doi: 10.5694/j.1326-5377.2005.tb07029.x] [Medline: 16138798]
6. Desai S, Jena AB. Do celebrity endorsements matter? Observational study of BRCA gene testing and mastectomy rates after Angelina Jolie's New York Times editorial. Br Med J 2016 Dec 14;355:i6357 [FREE Full text] [doi: 10.1136/bmj.i6357] [Medline: 27974323]
7. Evans DG, Barwell J, Eccles DM, Collins A, Izatt L, Jacobs C, et al. The Angelina Jolie effect: how high celebrity profile can have a major impact on provision of cancer related services. Breast Cancer Res 2014;16(5):442 [FREE Full text] [doi: 10.1186/s13058-014-0442-6] [Medline: 25510853]
8. Kluger J. The Angelina Effect: TIME's new cover image revealed. Time. 2013. URL: http://healthland.time.com/2013/05/15/the-angelina-effect-times-new-cover-image-revealed/ [accessed 2021-06-02]
9. Jolie A. My medical choice. New York Times. 2013. URL: http://i2-files.tv2net.dk/450/38119450-3a34246eaaf236d2118300c7d27c97b7.pdf [accessed 2021-06-02]
10. Noar SM, Willoughby JF, Myrick JG, Brown J. Public figure announcements about cancer and opportunities for cancer communication: a review and research agenda. Health Commun 2014;29(5):445-461. [doi: 10.1080/10410236.2013.764781] [Medline: 23845155]
11. Siegel RL, Miller KD, Sauer AG, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin 2020 May;70(3):145-164 [FREE Full text] [doi: 10.3322/caac.21601] [Medline: 32133645]
12. 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]
13. 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]
14. 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]
15. 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]
16. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):154-158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
17. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google trends research: systematic review. J Med Internet Res 2018 Nov 06;20(11):e270 [FREE Full text] [doi: 10.2196/jmir.9366] [Medline: 30401664]

18. Romero-Alvarez D, Parikh N, Osthus D, Martinez K, Generous N, Del Valle S, et al. Google Health Trends performance reflecting dengue incidence for the Brazilian states. BMC Infect Dis 2020 Mar 26;20(1):252 [FREE Full text] [doi: 10.1186/s12879-020-04957-0] [Medline: 32228508]
19. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PLoS One 2011 Aug 19;6(8):e23610 [FREE Full text] [doi: 10.1371/journal.pone.0023610] [Medline: 21886802]
20. Sadilek A, Hswen Y, Bavadekar S, Shekel T, Brownstein J, Gabrilovich E. Lymelight: forecasting Lyme disease risk using web search data. NPJ Digit Med 2020;3:16 [FREE Full text] [doi: 10.1038/s41746-020-0222-x] [Medline: 32047861]
21. Wehner MR, Nead KT, Linos E. Correlation among cancer incidence and mortality rates and internet searches in the United States. JAMA Dermatol 2017 Jun 28:911-914. [doi: 10.1001/jamadermatol.2017.1870] [Medline: 28658470]
22. Ayers JW, Westmaas JL, Leas EC, Benton A, Chen Y, Dredze M, et al. Leveraging big data to improve health awareness campaigns: a novel evaluation of the great American smokeout. JMIR Public Health Surveill 2016 Mar 31;2(1):e16 [FREE Full text] [doi: 10.2196/publichealth.5304] [Medline: 27227151]
23. Ayers JW, Althouse BM, Noar SM, Cohen JE. Do celebrity cancer diagnoses promote primary cancer prevention? Prev Med 2014 Jan;58:81-84. [doi: 10.1016/j.ypmed.2013.11.007] [Medline: 24252489]
24. Barros JM, Melia R, Francis K, Bogue J, O'Sullivan M, Young K, et al. The validity of Google trends search volumes for behavioral forecasting of national suicide rates in Ireland. Int J Environ Res Public Health 2019 Sep 02;16(17):1-18 [FREE Full text] [doi: 10.3390/ijerph16173201] [Medline: 31480718]
25. Leas EC, Nobles AL, Caputi TL, Dredze M, Smith DM, Ayers JW. Trends in internet searches for Cannabidiol (CBD) in the United States. JAMA Netw Open 2019 Oct 02;2(10):e1913853 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.13853] [Medline: 31642924]
26. Schootman M, Toor A, Cavazos-Rehg P, Jeffe DB, McQueen A, Eberth J, et al. The utility of Google Trends data to examine interest in cancer screening. BMJ Open 2015 Jun 08;5(6):e006678 [FREE Full text] [doi: 10.1136/bmjopen-2014-006678] [Medline: 26056120]
27. Kaminski M, Marlicz W, Koulaouzidis A. Googling on colonoscopy: a retrospective analysis of search engine statistics. Clin Exp Gastroenterol 2020 Oct;Volume 13:397-405. [doi: 10.2147/ceg.s266546]
28. Mukhija D, Venkatraman A, Nagpal SJ. Effectivity of awareness months in increasing internet search activity for top malignancies among women. JMIR Public Health Surveill 2017 Aug 21;3(3):e55 [FREE Full text] [doi: 10.2196/publichealth.7714] [Medline: 28827213]
29. Ayers JW, Althouse BM, Dredze M, Leas EC, Noar SM. News and internet searches about human immunodeficiency virus after Charlie Sheen's disclosure. JAMA Intern Med 2016 Apr;176(4):552-554. [doi: 10.1001/jamainternmed.2016.0003] [Medline: 26902971]
30. Ayers JW, Nobles AL, Dredze M. Media trends for the substance abuse and mental health services administration 800-662-HELP addiction treatment referral services after a celebrity overdose. JAMA Intern Med 2019 Mar 01;179(3):441-442 [FREE Full text] [doi: 10.1001/jamainternmed.2018.6562] [Medline: 30640378]
31. Hartwell M, Keener A, Coffey S, Chesher T, Torgerson T, Vassar M. Brief report: public awareness of asperger syndrome following Greta Thunberg appearances. J Autism Dev Disord 2021 Jun;51(6):2104-2108. [doi: 10.1007/s10803-020-04651-9] [Medline: 32812193]
32. Brigo F, Igwe SC, Nardone R, Lochner P, Tezzon F, Otte WM. Wikipedia and neurological disorders. J Clin Neurosci 2015 Jul;22(7):1170-1172. [doi: 10.1016/j.jocn.2015.02.006] [Medline: 25890773]
33. Bragazzi NL, Watad A, Brigo F, Adawi M, Amital H, Shoenfeld Y. Public health awareness of autoimmune diseases after the death of a celebrity. Clin Rheumatol 2016 Dec 20:1911-1917. [doi: 10.1007/s10067-016-3513-5] [Medline: 28000011]
34. Pagni F, Zàrate AF, Ren S. Neural invasion in pleomorphic adenoma of the salivary gland. Int J Surg Pathol 2010 Apr;18(2):151-152. [doi: 10.1177/1066896909349247] [Medline: 19948638]
35. Wikipedia pageviews analysis. Toolforge. URL: https://pageviews.toolforge.org [accessed 2021-06-02]
36. Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. J Stat Soft 2008;27(3):1-22. [doi: 10.18637/jss.v027.i03]
37. Diedenhofen B, Musch J. cocor: a comprehensive solution for the statistical comparison of correlations. PLoS One 2015 Apr 2;10(3):e0121945 [FREE Full text] [doi: 10.1371/journal.pone.0121945] [Medline: 25835001]
38. American Community Survey (ACS). United States Census Bureau. 2021. URL: https://www.census.gov/programs-surveys/acs [accessed 2021-06-12]
39. Carney K. Current census demographics by DMA. Cubit’s Blog. 2015. URL: https://blog.cubitplanning.com/2015/05/current-census-demographics-by-dma/ [accessed 2021-05-02]
40. Pantel HJ, Kleiman DA, Kuhnen AH, Marcello PW, Stafford C, Ricciardi R. Has national colorectal cancer awareness month increased endoscopy screening rates and public interest in colorectal cancer? Surg Endosc 2021 Jan;35(1):398-405. [doi: 10.1007/s00464-020-07413-x] [Medline: 32016518]
41. Plackett R, Kaushal A, Kassianos AP, Cross A, Lewins D, Sheringham J, et al. Use of social media to promote cancer screening and early diagnosis: scoping review. J Med Internet Res 2020 Nov 09;22(11):e21582 [FREE Full text] [doi: 10.2196/21582] [Medline: 33164907]

42. Bakouny Z, Paciotti M, Schmidt AL, Lipsitz SR, Choueiri TK, Trinh Q. Cancer screening tests and cancer diagnoses during the COVID-19 pandemic. JAMA Oncol 2021 Mar 01;7(3):458-460. [doi: 10.1001/jamaoncol.2020.7600] [Medline: 33443549]

Abbreviations

BRCA1: breast cancer gene 1
DMA: designated market area
GT: Google Trends
RSV: relative search volume

Edited by R Kukafka; submitted 08.01.21; peer-reviewed by A Mavragani, T Caputi, M Kaminski; comments to author 23.01.21; revised version received 13.05.21; accepted 24.05.21; published 15.06.21.

Please cite as:
Naik H, Johnson MDD, Johnson MR
Internet Interest in Colon Cancer Following the Death of Chadwick Boseman: Infoveillance Study
J Med Internet Res 2021;23(6):e27052
URL: https://www.jmir.org/2021/6/e27052
doi:10.2196/27052
PMID:34128824

Original Paper

Biannual Differences in Interest Peaks for Web Inquiries Into Ear Pain and Ear Drops: Infodemiology Study

Faris F Brkic[1], MD; Gerold Besser[1], PhD, MD; Martin Schally[1]; Elisabeth M Schmid[1]; Thomas Parzefall[1], MD, PhD; Dominik Riss[1], MD; David T Liu[1], MD

[1] Department of Otorhinolaryngology, Head and Neck Surgery, Medical University of Vienna, Vienna, Austria

Corresponding Author:
David T Liu, MD
Department of Otorhinolaryngology, Head and Neck Surgery
Medical University of Vienna
Währinger Gürtel 18-20
Vienna, 1090
Austria
Phone: 43 1 404 002 0820
Email: david.liu@meduniwien.ac.at

Abstract

Background: The data retrieved with the online search engine, Google Trends, can summarize internet inquiries into specified search terms. This engine may be used for analyzing inquiry peaks for different medical conditions and symptoms.

Objective: The aim of this study was to analyze World Wide Web interest peaks for “ear pain,” “ear infection,” and “ear drops.”

Methods: We used Google Trends to assess the public online interest for search terms “ear pain,” “ear infection,” and “ear drops” in 5 English and non–English-speaking countries from both hemispheres based on time series data. We performed our analysis for the time frame between January 1, 2004, and December 31, 2019. First, we assessed whether our search terms were most relevant to the topics of ear pain, ear infection, and ear drops. We then tested the reliability of Google Trends time series data using the intraclass correlation coefficient. In a second step, we computed univariate time series plots to depict peaks in web-based interest. In the last step, we used the cosinor analysis to test the statistical significance of seasonal interest peaks.

Results: In the first part of the study, it was revealed that “ear infection,” “ear pain,” and “ear drops” were the most relevant search terms in the noted time frame. Next, the intraclass correlation analysis showed a moderate to excellent reliability for all 5 countries’ 3 primary search terms. The subsequent analysis revealed winter interest peaks for “ear infection” and “ear pain”. On the other hand, the World Wide Web search for “ear drops” peaked annually during the summer months. All peaks were statistically significant as revealed by the cosinor model (all *P* values <.001).

Conclusions: It can be concluded that individuals affected by otitis media or externa, possibly the majority, look for medical information online. Therefore, there is a need for accurate and easily accessible information on these conditions in the World Wide Web, particularly on differentiating signs and therapy options. Meeting this need may facilitate timely diagnosis, proper therapy, and eventual circumvention of potentially life-threatening complications.

(J Med Internet Res 2021;23(6):e28328) doi:10.2196/28328

KEYWORDS

otitis media; otitis externa; otalgia; Google Trends; infodemiology; infoveillance; infodemic; social listening

Introduction

Google Search (Google LLC) is the most commonly used internet search engine, and almost two-thirds of all daily online inquiries are performed using this platform. [1] Google Trends (GT) delivers information on geographical and temporal patterns of search volumes for inquiries performed with Google. [2] Notably, several studies have already used GT to analyze web user searches for different otolaryngological symptoms, conditions, and diseases. [3-6] Furthermore, conditions in other medical disciplines have been assessed, including in cardiology [7,8], urology [9], and orthopedic surgery [10]. In general, the internet has become an appealing source of health care information, particularly for younger people, who are more amenable to new and innovative approaches for gaining novel information in an online environment [11]. Therefore, internet search volumes for certain medical conditions could potentially follow their annual incidence peaks.

Acute otitis media (OM) and otitis externa (OE) are common conditions in otolaryngology and represent the middle ear and external auditory canal's inflammation, respectively [12,13]. OM is mostly a result of a dysfunction of the auditive tube due to upper airway tract infections [12]. New diagnostic guidelines recommend combining medical history, clinical symptoms, and strict otoscopic criteria to determine this diagnosis [14]. These include moderate or severe tympanic membrane bulging combined with otorrhea or mild bulging combined with recent onset of ear pain. OE's incidence is comparable to that of OM and affects up to 10% of people at some point in life. [13] Some common symptoms tend to overlap with those of OM, such as ear pain or otorrhea. Nevertheless, some critical differentiation signs exist, such as ear itching and tenderness of the tragus or pinna, which are OE-specific symptoms [15]. Based on the many similarities between these two conditions, it may be assumed that individuals tend to misdiagnose or incorrectly treat themselves by referring to information available online.

It has been noted that OM occurs mostly in the winter months, which correlates with the incidence rates of upper airway infection [12,16-19]. In contrast, OE incidence rises significantly during the warmer summer months, mostly due to increased humidity, sweating, and swimming (ie, water exposure to the outer ear canal [15,20,21]). Interestingly, a study on seasonal peaks of acute OM incidence revealed peaks in the winter and summer months [22].

To the best of our knowledge, only one study group has assessed online user behavior regarding otologic conditions. Specifically, they correlated GT search for "ear drops" with Medicaid prescription frequency for ototopical agents [23]. However, to date, no study has analyzed World Wide Web public inquiries into other OM- and OE-related terms (such as ear pain or ear infection). Therefore, our study aimed to analyze the web-based interest into ear pain, ear infections, and the associated treatment options (ear drops). This study's results may help gain novel insights into the temporal frequency of internet searches into ear pain–related search terms and clarify the temporal dynamics of the affected individuals who search for symptoms or treatment options regarding these medical conditions.

Methods

GT was used to explore the online search interest for "ear pain" and related terms entered into the Google search engine from various countries worldwide. The relative search volume (RSV) shows user interest in specific search terms. It ranges from 0 to 100 (higher interest correlates with higher score). The normalization steps are described elsewhere [24]. We explored RSV for the following 3 search terms (and the country-specific translation) related to ear infections: "ear drops" ("Ohrentropfen"), "ear pain" ("Ohrenschmerzen"), and "ear infection" ("Ohr Entzündung"). In order to cover both hemispheres and 2 languages (English and German), we assessed internet-based inquiries in the following countries: Australia, Canada, Germany, the United States, and the United Kingdom. The data were retrieved in the "Health" category for the time frame between January 1, 2004, and December 31, 2019.

The 3 preliminary search terms were entered on August 18, 2020, and the function "Related (Top)" was used to show search terms related to the preliminary search terms in every analyzed country. We ensured that we analyzed 3 search terms with the highest RSV in the context of "ear pain," "ear infection," and "ear drops" by using the GT function "Comparison." We therefore compared the RSV of all related search terms against the 3 preliminary search terms. The most relevant search terms associated with each of the 3 preliminary search terms were used for further data acquisition and analyses. Furthermore, we also added the search terms "otitis" and "otitis media" ("Gehörgangsentzündung" and "Mittelohrentzündung", respectively, for Germany) to cover the full spectrum of ear pain–related search terms.

We then entered each of the aforementioned 5 ear pain–related search terms (ie, "ear drops," "ear pain," "ear infection," "otitis," and "otitis media") on April 25, 2021, and used the function "Worldwide" to assess regional differences in country-specific RSV. We specified our searches for the "Health" category and the time frame between January 1, 2004, and December 31, 2019. Finally, a reliability analysis was performed by downloading selected search terms for 10 consecutive days, starting from August 18, 2020. It has previously been reported that slightly different results are shown if GT data are downloaded on different days [3,16].

The analysis was performed with R software version 3.5.1 (R Foundation for Statistical Computing) with the "season" and "psych" packages. The visualization of these data was performed with Prism 9.0.0 software (GraphPad). The intraclass correlation coefficient (ICC; 2-way random model) was used for examination of GT data reliability. Poor reliability was defined as an ICC lower than 0.4, moderate reliability as an ICC higher than or equal to 0.4 and lower than 0.75, good reliability as an ICC higher than or equal to 0.75 and lower than 0.9, and excellent reliability as an ICC greater than or equal to 0.9. The cosinor regression model was used to detect seasonality in time series data. The exact model is described in detail elsewhere [25]. In short, the cosinor model is a parametric model that captures seasonal patterns using a sinusoid. We fitted an annual seasonal pattern to our time series data and therefore defined the annual seasonal cycle as 12 (months). As the cosinor model assumes the sinusoid to be symmetric and stationary, a peak (P, peak) and a nadir point (L, low point, defined as peak month +/–6 months) are defined once per year. The cosinor analysis also computes an amplitude (size), which represents the magnitude of the seasonal effect. As the sinusoid is described by both a sine and a cosine term, statistical significance can be tested as part of a generalized linear model. We set the *P* value to .03 to adjust for multiple comparisons (sine and cosine term).

The visualization of worldwide country-specific differences in RSV for our 5 ear pain–related search terms (mentioned previously) was performed with Python 3.9.4 [26] in combination with the libraries NumPy [27], Pandas [28], Matplotlib [29], and Geopandas [30].

For the current study, only publicly available and nonidentifiable data were explored. Therefore, according to the guidelines of the institutional review board of the Medical University of Vienna, study approval was not needed.

Results

Search Terms Related to Ear Infection, Ear Pain, and Ear Drops

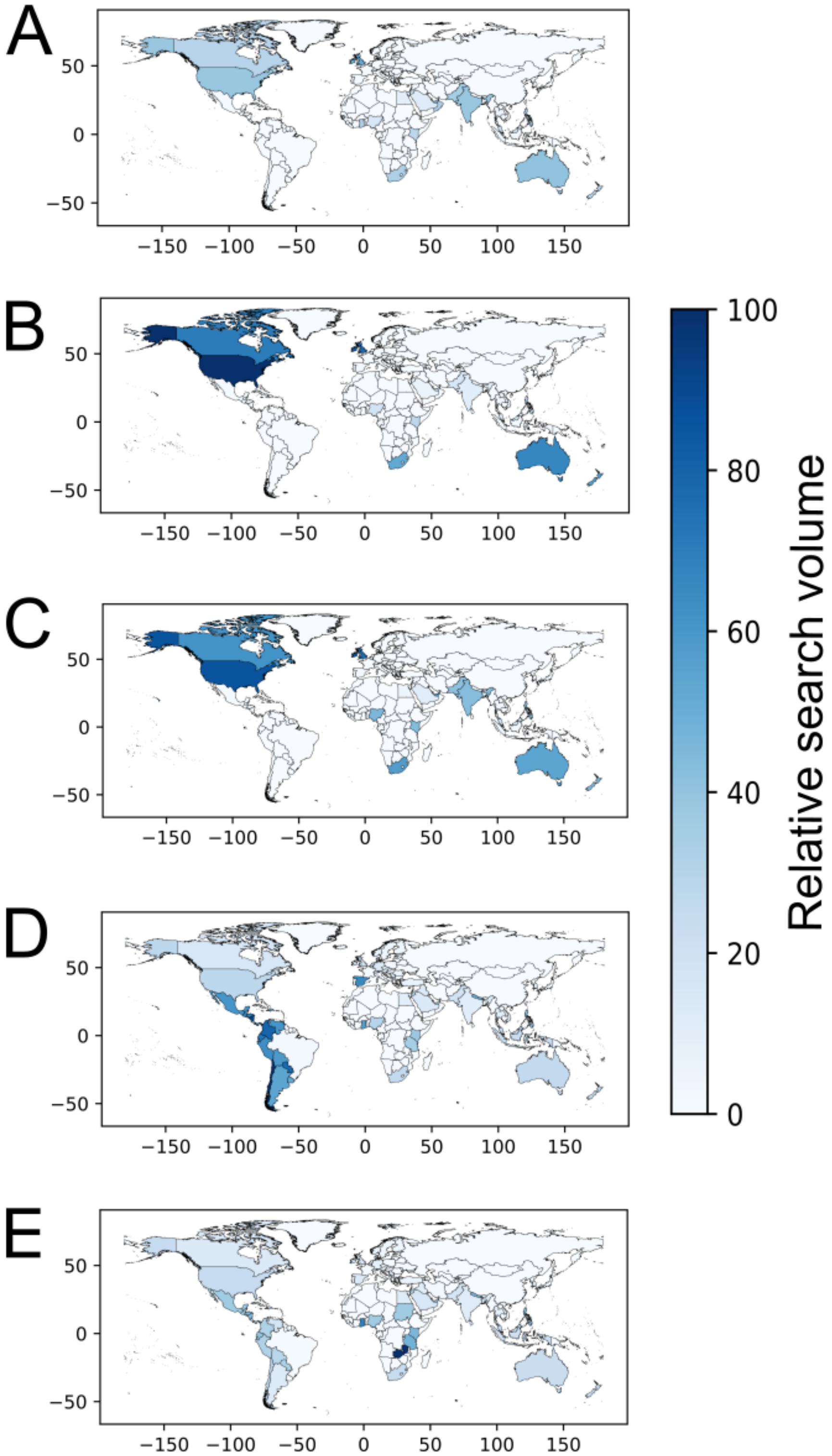
As we wanted to perform the analysis only using search terms with the highest RSV, our first step was to assess if preliminary search terms were in fact the most searched in regard to the following 3 terms: (1) “ear pain,” (2) “ear infection,” and (3) “ear drops.” We therefore entered each of the 3 primary search terms into GT for the time frame between January 1, 2004, and December 31, 2019. The search was performed using the “Health” category. Furthermore, the GT function “Related queries (Top)” was used to identify search terms that users entered on GT following the specified one. This step was followed by comparing the RSV of each related search term with each of the 3 preliminary search terms to evaluate whether the latter terms had the highest RSV within their respective category. It was indeed shown that our 3 preliminary search terms (“ear pain,” “ear infection,” and “ear drops”) represented the 3 most relevant search terms within their respective category in all English-speaking countries. Only the search term “Mittelohrentzündung” (middle ear infection) had a higher RSV compared to “Ohr Entzündung” (ear infection) in Germany (Tables S1-S5, Multimedia Appendix 1). To cover the full spectrum of ear pain–related search terms, we also added “otitis” and “otitis media” (“Gehörgangsentzündung” and “Mittelohrentzündung” for Germany) for further data analyses.

Regional Differences in Relative Search Volume for Ear Infection–Related Search Terms

In the next step, we assessed country-specific differences in RSV for our 5 ear infection–related search terms: “ear pain,” “ear infection,” “ear drops,” “otitis,” and “otitis media.” We used the GT function “Worldwide” to compare the country-specific RSVs.

The graphical analysis revealed regional differences in RSV for each of the 5 ear infection–related search terms. “Ear infection” and “ear pain” were searched more often in English-speaking countries from the Northern Hemisphere, while “otitis” and “otitis media” were searched more often in countries from the Southern Hemisphere (Figure 1).

Figure 1. World maps showing country-specific variations in relative search volume for (A) “ear drops,” (B) “ear infection,” (C) “ear pain,” (D) “otitis,” and (E) “otitis media.” The color legend represents the relative search volume. The x-axis and the y-axis represent the longitude and latitude, respectively. The maps have been created with the GeoPandas library [30].



Reliability of Ear Infection–Related Search Terms

We performed the reliability analysis using the ICC for GT time series data of our 3 final search terms extracted on 10 consecutive days, starting on August 18, 2020.

The 3 primary search terms showed good to excellent reliability in English-speaking countries, with correlation coefficients ranging between 0.78 and 0.97 for “ear drops,” between 0.87 and 0.99 for “ear infection,” and between 0.89 to 0.99 for “ear pain.” Furthermore, reliability analyses in Germany also revealed excellent correlation coefficients for 2 of the 3 primary search terms: ICC of 0.91 for “Ohrentropfen” (ear drops), ICC of 0.47 for “Ohr Entzündung” (ear infection), and ICC of 0.98 for “Ohrenschmerzen” (ear pain; Table 1). The analysis revealed less reliable results for our 2 additional search terms. The ICC ranged between 0.36 and 0.86 for “otitis” and between 0.37 to 0.74 for “otitis media” in the English-speaking countries. Further correlation analyses revealed an ICC of 0.93 for “Mittelohrentzündung” and an ICC of 0.75 for “Gehörgangsentzündung” in Germany.

Table 1. Reliability of single and averaged time series data on "ear drops," "ear infection," "ear pain," "otitis," and "otitis media" in all countries.

Search term by country	ICC[a]	Lower bound[b]	Upper bound[b]	*F* test[c]	Df1[d]	Df2[e]	*P* value
Australia							
Ear drops							
Single[f]	0.78	0.74	0.81	35.96	191	1719	<.001
Average[g]	0.97	0.97	0.98	35.96	191	1719	<.001
Ear infection							
Single	0.87	0.84	0.89	71.81	191	1719	<.001
Average	0.99	0.98	0.99	71.81	191	1719	<.001
Ear pain							
Single	0.89	0.87	0.91	88.27	191	1719	<.001
Average	0.99	0.99	1.00	88.27	191	1719	<.001
Otitis							
Single	0.36	0.27	0.44	8.90	191	1719	<.001
Average	0.85	0.79	0.89	8.90	191	1719	<.001
Otitis media							
Single	0.46	0.39	0.54	11.70	191	1719	<.001
Average	0.90	0.87	0.92	11.70	191	1719	<.001
Canada							
Ear drops							
Single	0.83	0.80	0.86	52.04	191	1719	<.001
Average	0.98	0.98	0.98	52.04	191	1719	<.001
Ear infection							
Single	0.87	0.84	0.89	74.41	191	1719	<.001
Average	1.99	0.98	0.99	74.41	191	1719	<.001
Ear pain							
Single	0.92	0.89	0.93	123.40	191	1719	<.001
Average	0.99	0.99	0.99	123.40	191	1719	<.001
Otitis							
Single	0.86	0.83	0.89	73.57	191	1719	<.001
Average	0.98	0.98	0.99	73.57	191	1719	<.001
Otitis media							
Single	0.47	0.41	0.53	10.57	191	1719	<.001
Average	0.90	0.87	0.92	10.57	191	1719	<.001
United Kingdom							
Ear drops							
Single	0.97	0.96	0.97	300.23	191	1719	<.001
Average	1.00	1.00	1.00	300.23	191	1719	<.001
Ear infection							
Single	0.98	0.98	0.99	765.13	191	1719	<.001
Average	1.00	1.00	1.00	765.13	191	1719	<.001
Ear pain							
Single	0.99	0.98	0.99	851.84	191	1719	<.001

Search term by country	ICC[a]	Lower bound[b]	Upper bound[b]	*F* test[c]	Df1[d]	Df2[e]	*P* value
Average	1.00	1.00	1.00	851.84	191	1719	<.001
Otitis							
Single	0.33	0.18	0.47	17.05	191	1719	<.001
Average	0.83	0.68	0.90	17.05	191	1719	<.001
Otitis media							
Single	0.37	0.24	0.48	13.00	191	1719	<.001
Average	0.85	0.76	0.90	13.00	191	1719	<.001
United States							
Ear drops							
Single	0.97	0.97	0.98	404.42	191	1719	<.001
Average	1.00	1.00	1.00	404.42	191	1719	<.001
Ear infection							
Single	0.99	0.99	0.99	1096.64	191	1719	<.001
Average	1.00	1.00	1.00	1096.64	191	1719	<.001
Ear pain							
Single	0.99	0.99	0.99	1523.05	191	1719	<.001
Average	1.00	1.00	1.00	1523.05	191	1719	<.001
Otitis							
Single	0.51	0.36	0.64	25.50	191	1719	<.001
Average	0.91	0.85	0.95	25.50	191	1719	<.001
Otitis media							
Single	0.74	0.70	0.79	32.61	191	1719	<.001
Average	0.97	0.96	0.97	32.61	191	1719	<.001
Germany							
Mittelohrentzündung							
Single	0.93	0.91	0.94	132.45	191	1719	<.001
Average	0.99	0.99	0.99	132.45	191	1719	<.001
Ohrenschmerzen							
Single	0.98	0.98	0.99	691.66	191	1719	<.001
Average	1.00	1.00	1.00	691.66	191	1719	<.001
Ohrentropfen							
Single	0.91	0.89	0.92	98.68	191	1719	<.001
Average	0.99	0.99	0.99	98.68	191	1719	<.001
Ohr Entzündung							
Single	0.47	0.38	0.55	12.93	191	1719	<.001
Average	0.90	0.86	0.92	12.93	191	1719	<.001
Gehörgangsentzündung							
Single	0.75	0.71	0.79	32.38	191	1719	<.001

Search term by country	ICC[a]	Lower bound[b]	Upper bound[b]	*F* test[c]	Df1[d]	Df2[e]	*P* value
Average	0.97	0.96	0.98	32.38	191	1719	<.001

[a] ICC: intraclass correlation coefficient.

[b] lower and upper bound = 95% confidence interval of the intraclass correlation coefficient.

[c] *F* test for significance of the correlation coefficient.

[d] Df1: numerator degrees of freedom.

[e] Df2: denominator degrees of freedom.

[f] Single: single time series data.

[g] Average: averaged time series data.

Biannual Differences in Web-Based Interest Peaks for Our Ear-Related Search Terms

Next, we performed a time series analysis to depict the seasonal variations graphically and statistically. The data sets extracted on the first day of data retrieval (August 18, 2020) were used for this analysis. We computed univariate time series plots to analyze monthly changes in RSV graphically (Figures 2-4). Subsequently, the cosinor model was used to analyze the time series data to depict peaks and nadirs in web-based interest for each of the 3 final search terms.

The graphical analysis revealed interest peaks in the summer months for “ear drops” in English- and non–English-speaking countries from both hemispheres (Figure 5). In contrast, peaks in web-based interest for “ear infection” and “ear pain” were apparent in the winter months. Subsequent cosinor analyses revealed all of these peaks to be significant (all *P* values <.001) (Table 2).

Figure 2. Univariate time series plot showing monthly changes in relative search volume for (A) “ear infection,” (B) “ear drops,” and (C) “ear pain” in the United States between January 2004 and December 2019. The thick vertical orange lines mark January 1, while the thin vertical green lines mark July 1. The black points represent the relative search volume of each of the 12 months.

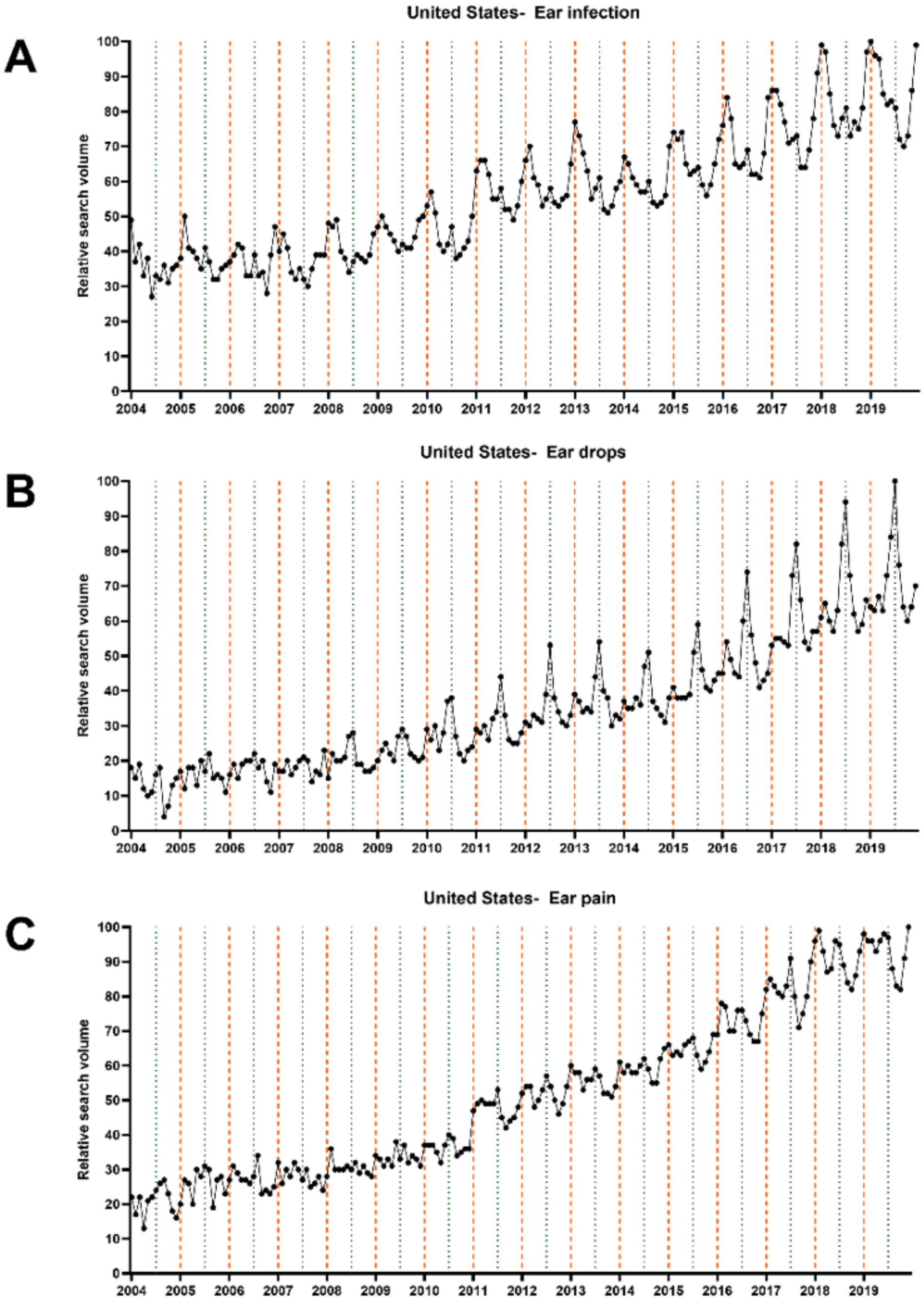


Figure 3. Univariate time series plot showing monthly changes in relative search volume for (A) “ear infection,” (B) “ear drops,” and (C) “ear pain” in Australia between January 2004 and December 2019. The thick vertical orange lines mark January 1, while the thin vertical green lines mark July 1. The black points represent the relative search volume of each of the 12 months.

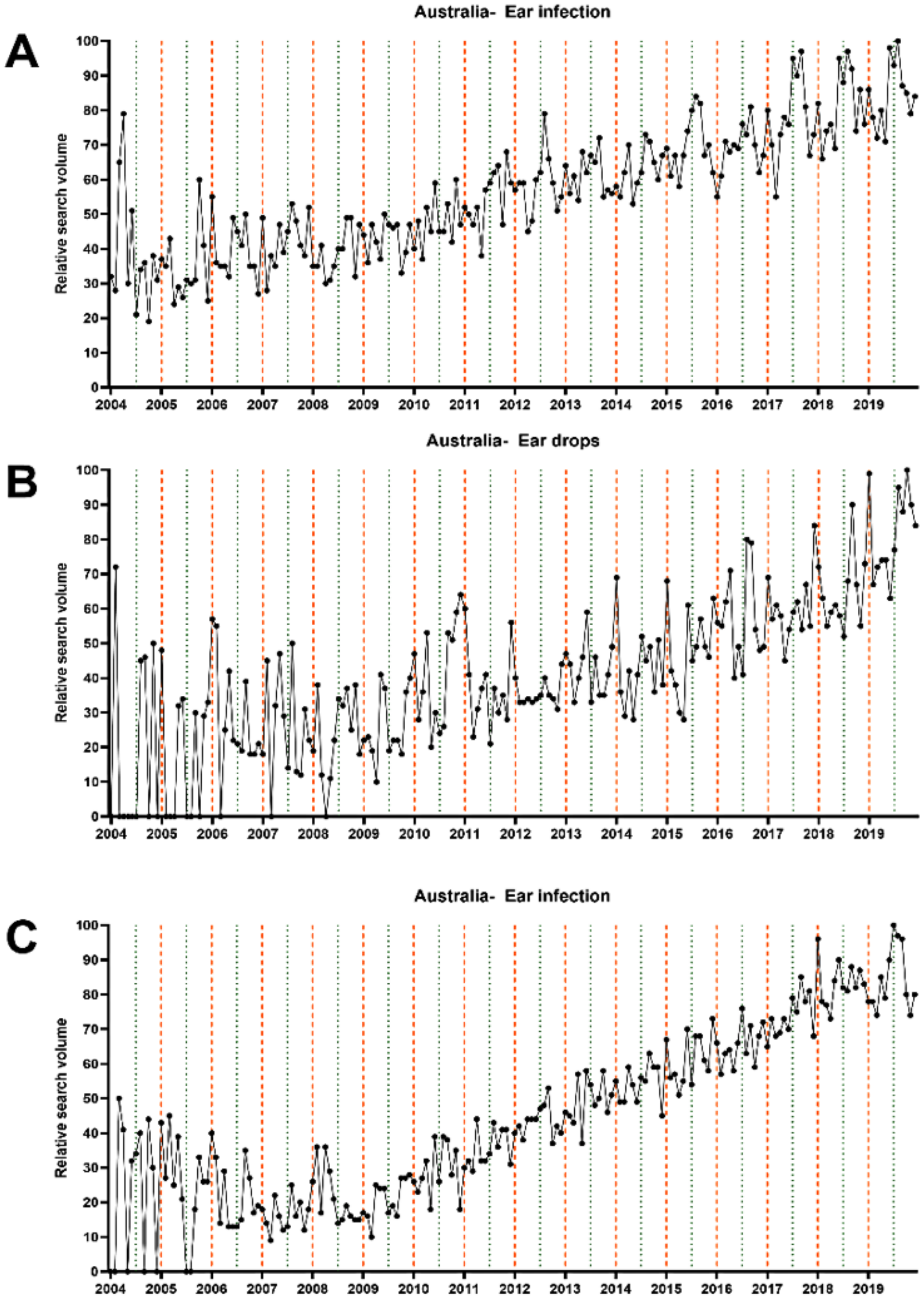


Figure 4. Univariate time series plot showing monthly changes in relative search volume for (A) “ear infection,” (B) “ear drops,” and (C) “ear pain” in Germany between January 2004 and December 2019. The thick vertical orange lines mark January 1, while the thin vertical green lines mark July 1. The black points represent the relative search volume of each of the 12 months.

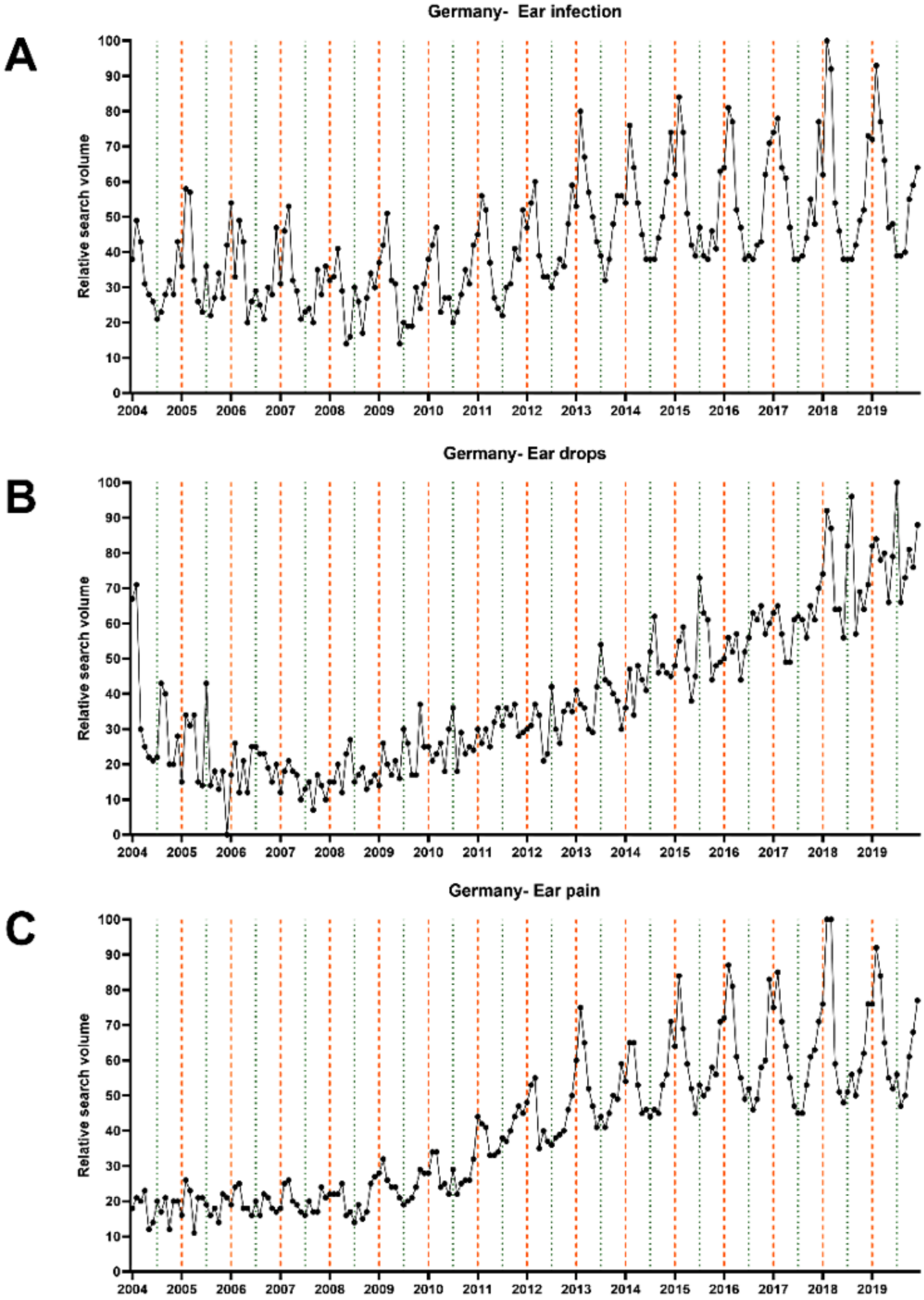


Figure 5. Cosinor analysis plots showing monthly variations in relative search volume for "ear infection," "ear pain," and "ear drops" (from left to right) in (A) the United States, (B) Germany, and (C) Australia. The black dots mark the monthly mean relative search volume, while the error bars mark the SE. The numbers on the x-axis represent the corresponding months (ie, 1 = January, 2 = February).

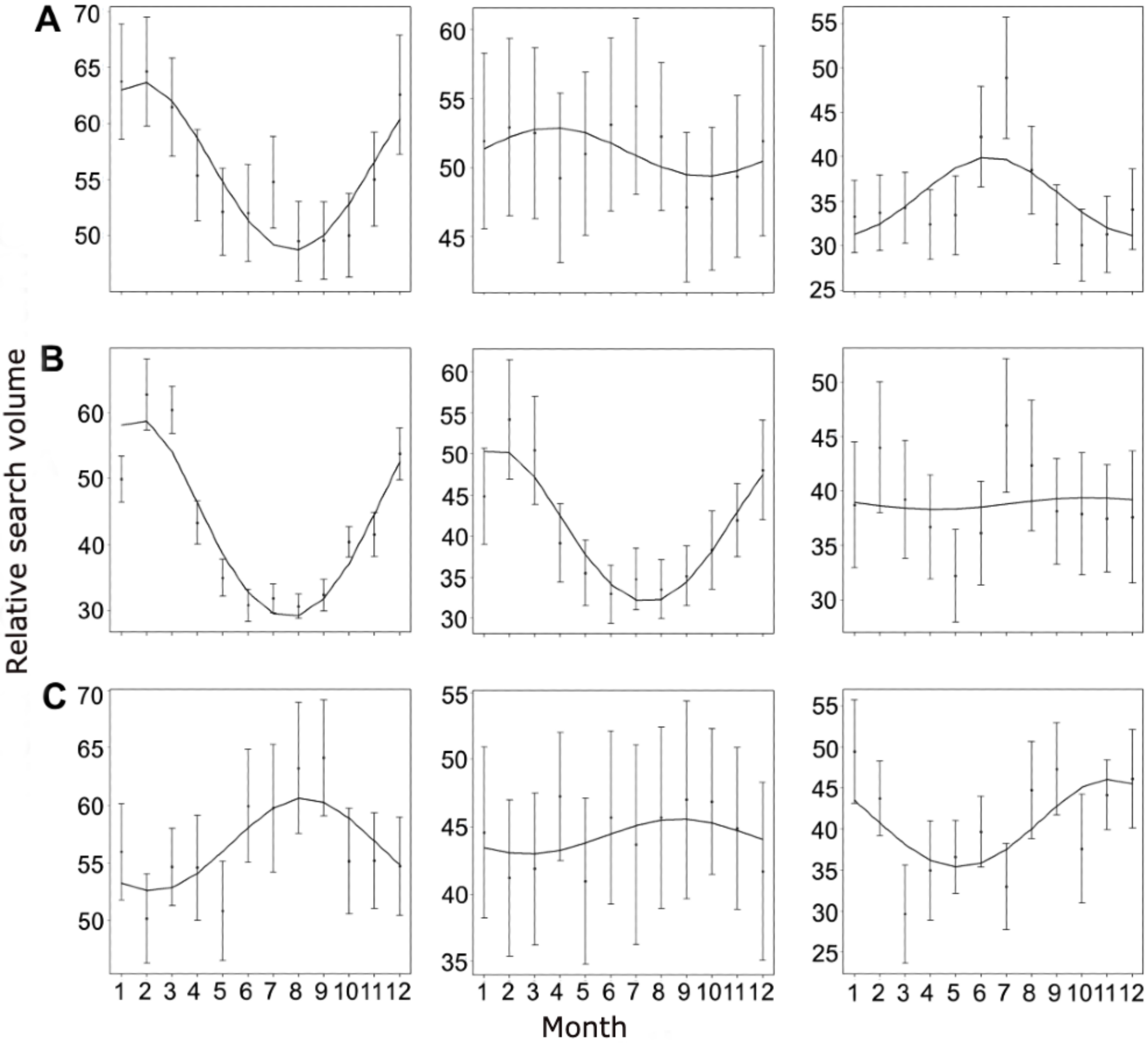


Table 2. Cosinor analysis on seasonality of “ear drops,” “ear pain,” “ear infection,” “otitis,” and “otitis media”.

Term by country	Amplitude	Peak[a]	Nadir[a]	SE	*P* value
Australia					
Ear drops	5.73	11.1	5.1	0.016	<.001
Ear pain	1.32	8.7	2.7	0.015	<.001
Ear infection	4.20	8.2	2.2	0.014	<.001
Otitis	3.67	6.0	12.0	0.022	<.001
Otitis media	3.82	6.9	12.9	0.023	.002
Canada					
Ear drops	0.91	7.3	1.3	0.018	<.001
Ear pain	1.68	2.1	8.1	0.015	<.001
Ear infection	9.77	1.5	7.5	0.013	<.001
Otitis	11.13	1.2	7.2	0.015	<.001
Otitis media	6.44	2.0	8.0	0.018	<.001
United Kingdom					
Ear drops	2.63	7.6	1.6	0.016	<.001
Ear pain	2.31	12.3	6.3	0.017	<.001
Ear infection	4.59	12.5	6.5	0.016	<.001
Otitis	3.90	2.0	8.0	0.015	<.001
Otitis media	7.76	1.4	7.4	0.018	<.001
United States					
Ear drops	4.74	6.4	12.4	0.017	<.001
Ear pain	1.79	3.7	9.7	0.014	<.001
Ear infection	8.03	1.8	7.8	0.014	<.001
Otitis	3.18	1.7	7.7	0.011	<.001
Otitis media	8.07	1.7	7.7	0.012	<.001
Germany					
Mittelohrentzündung	17.78	1.6	7.6	0.016	<.001
Ohrenschmerzen	10.36	1.5	7.5	0.016	<.001
Ohrentropfen	0.55	10.3	4.3	0.016	<.001
Ohr entzündung	1.98	8.9	2.9	0.025	<.001
Gehörgangsentzündung	12.72	8.0	2.0	0.018	<.001

[a] The values in this column correspond to the respective month (eg, 1=January, 2=February).

Discussion

The current study used GT and revealed winter and summer peaks in online search volumes for OE- and OM-related search terms, which corresponded with the winter and summer incidence rise of these conditions. In particular, highly reliable results revealed that the search volume for “ear drops” peaked in the summer months, potentially reflecting the interest in OE at this time. On the contrary, the annual increase in search frequency for terms “ear pain” and “ear infection” (hypothetically reflecting both OE and OM) was observed during the winter months. This discrepancy between online search peaks could indicate a potential mistake in treatment by self-diagnosis or self-treatment of either of these two conditions by patients searching for therapy options online. As noted, one study already assessed this problem by correlating Google search frequency for ear drops and Medicaid prescriptions for ototopical agents [23]. That study, however, included some methodological limitations: no further GT analysis regarding related terms or symptoms of OM or OE was performed, the reliability of the GT data was not analyzed, and the online search frequency for noted terms was assessed for only 1 country. Therefore, the significance of our study is mainly its highly reliable GT data, reproducibility in different countries from both hemispheres, and significant interest peaks revealed by the cosinor analysis for all 3 search terms.

As noted, GT has often been used to analyze user-associated online behavior in regard to different otorhinolaryngologic conditions [3-6,16]. Furthermore, public online interest for certain oral, maxillofacial, and dentistry problems has been analyzed by several authors [31-33]. Moreover, this engine was used for assessing public inquiries into seasonal influenza [34,35]. Even the currently prevailing and life-changing COVID-19 pandemic was analyzed [36,37], and authors have also discussed using GT to predict COVID-19 outbreaks [38,39].

There are several critical similarities between OM and OE regarding symptoms and therapy options. The overlapping symptoms include ear pain and otorrhea [14,15], while specific OE signs include an itchy ear and tragus pain [15]. Therapy regimens for both conditions include analgesics, but the indication for antibiotic therapy varies significantly. Although bacteria can be isolated from the middle ear in 50% to 90% of cases [40,41], not all OM patients require antibiotic therapy at initial presentation. Clinical observation and analgesic therapy are the first-line treatment approaches in adult patients and children over 6 years with mild symptoms. However, either a clinical follow-up within 2 to 3 days should be scheduled, or a backup antibiotic should be prescribed and used in cases where symptoms persist [14,42]. Regarding OE, therapy options include ear cleaning, analgesics, and topical treatment (antibiotics, steroids, antiseptics, antifungals, or combinations) [43]. Oral antibiotics are not indicated for treatment of OE, as they prolong time to clinical cure and are not associated with better outcomes as compared with a topical agent used alone in uncomplicated cases [20,44]. Additionally, the overuse of systemic antibiotics contributes to the global problem of antibiotic resistance [16]. In summary, systemic or oral antibiotic is only required in severe and persistent OM, while OE can be treated mostly with topical therapy using ear drops.

Incompletely treated or untreated OM can result in complications, such as mastoiditis, facial nerve paresis or palsy, or labyrinthitis. Therefore, symptoms of OM warrant further diagnosis and therapeutic approaches [45]. Furthermore, these can occur in cases with inadequately treated OM (eg, in cases of antibiotic-resistant bacteria) [45]. The most typical symptoms of acute mastoiditis are a retroauricular swelling and a protruded pinna. Other clinical signs include retroauricular erythema and pain, or an abscess of the external auditory canal [46,47]. Timely recognition of these signs and a prompt referral to an otolaryngology department are crucial in treating OM complications. The next steps include intravenous antibiotic therapy and a computer tomography scan in cases that do not respond to therapy [48]. Surgical treatment such as mastoidectomy remains the most efficient therapy option for patients with intracranial complications or mastoid abscess [49]. Malignant (necrotizing) OE is a potentially fatal complication of acute OE and affects older adults, immunocompromised individuals, and patients with undertreated diabetes mellitus [50]. The most common signs are otalgia, subacute hearing loss, and intensive otorrhea [50]. The infection can extend to the mastoid and the skull base and can potentially result in facial nerve palsy, venous sinus thrombosis, osteomyelitis, or even death [51]. Although patients potentially self-manage these conditions by looking for information online, timely referral to the hospital appears crucial, as a definitive diagnosis can only be made by a medical professional. Not postponing a doctor's appointment can lead to proper treatment and circumvention of potentially life-threatening complications.

OM affects all age groups, but about 80% of children have at least 1 acute OM episode before school age [12]. However, a plethora of pediatric patients does not necessarily warrant antibiotic therapy. Systemic antibiotics should be routinely indicated only in children above the age of 6 months with severe symptoms and in children older than 2 years with bilateral acute OM [14]. Observation with scheduled clinical follow-up is recommended in children 2 to 12 years old with nonsevere symptoms [14]. Furthermore, OM with effusion is often misdiagnosed as acute OM and overtreated with antibiotics [52]. Overtreatment is also a common problem affecting patients with OE. One study group noted that about 20% to 40% of patients with an acute OE are primarily and unnecessarily treated with a systemic antibiotic [53] even though systemic antibiotics do not necessarily produce better clinical outcomes [20]. In summary, the timely differentiation between OM and OE is crucial, as it can result in proper therapy, reduction of overtreatment, and circumvention of possibly life-threatening complications.

The current study faces limitations associated with its infodemiological design. Potential standard limitations of infodemiological studies could influence the interpretation of our results. As we used a single search engine to retrieve all data, selection bias is a potential limitation. However, about two-thirds of daily internet searches are performed using Google web search, which minimizes this risk [1]. A further selection bias is reflected by the fact that people from higher-income areas have access and tend to use the internet for any kind of information, particularly regarding certain medical conditions [54]. Furthermore, with GT, the data on the age or gender of users who search for different terms cannot be assessed. This limitation could be a confounding factor, as younger people are more likely to use online search engines to gather information on medical conditions [55]. On the other hand, studies designed around using infodemiological methods can arguably be more extensive, detailed, and real time than epidemiological studies. Therefore, with these methods, the information retrieval and the efficacy of the research can be improved. Furthermore, it can only be assumed that RSV varies only slightly across individual areas within each country. We included larger countries in our analysis. Thus, our results may only represent those regions with more inhabitants (and therefore higher RSV). Future studies will reveal whether regional differences have a significant influence on online search frequency for different medical conditions.

In conclusion, we observed biannual (summer and winter) peaks in searches for otitis externa and media and related terms, which correlated with the respective annual incidence increase of these two conditions. These findings thus underline the necessity for accurate and easily accessible medical information on the internet, particularly for diagnosis, appropriate therapy options, and differentiating between OE and OM. This type of information may reduce overtreatment with antibiotics in OE cases and mitigate the global problem of antibiotic resistance.

Finally, prompt and early detection of potentially life-threatening complications and subsequent further diagnosis and therapy could be facilitated.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Results from relative search volume comparison between primary and related search terms in Australia.

[DOCX File , 70 KB - jmir_v23i6e28328_app1.docx]

References

1. Browser market share. Market Share Statistics for Internet Technologies. URL: https://netmarketshare.com [accessed 2020-12-20]
2. Pier MM, Pasick LJ, Benito DA, Alnouri G, Sataloff RT. Otolaryngology-related Google Search trends during the COVID-19 pandemic. Am J Otolaryngol 2020;41(6):102615 [FREE Full text] [doi: 10.1016/j.amjoto.2020.102615] [Medline: 32659612]
3. Liu DT, Besser G, Leonhard M, Bartosik TJ, Parzefall T, Brkic FF, et al. Seasonal variations in public inquiries into laryngitis: an infodemiology study. J Voice 2020 May 19:1-8 [FREE Full text] [doi: 10.1016/j.jvoice.2020.04.018] [Medline: 32439216]
4. Liu DT, Besser G, Parzefall T, Riss D, Mueller CA. Winter peaks in web-based public inquiry into epistaxis. Eur Arch Otorhinolaryngol 2020 Jul;277(7):1977-1985 [FREE Full text] [doi: 10.1007/s00405-020-05915-x] [Medline: 32180015]
5. Plante DT, Ingram DG. Seasonal trends in tinnitus symptomatology: evidence from Internet search engine query data. Eur Arch Otorhinolaryngol 2015 Oct;272(10):2807-2813. [doi: 10.1007/s00405-014-3287-9] [Medline: 25234771]
6. Faoury M, Upile T, Patel N. Using Google Trends to understand information-seeking behaviour about throat cancer. J. Laryngol. Otol 2019 Jul 08;133(7):610-614. [doi: 10.1017/s0022215119001348]
7. Trohman RG, Sharma PS, McAninch EA, Bianco AC. Amiodarone and thyroid physiology, pathophysiology, diagnosis and management. Trends Cardiovasc Med 2019 Jul;29(5):285-295 [FREE Full text] [doi: 10.1016/j.tcm.2018.09.005] [Medline: 30309693]
8. Zuin M, Rigatelli G, Ronco F. Worldwide and European interest in the MitraClip. Journal of Cardiovascular Medicine 2020;21(3):246-249. [doi: 10.2459/jcm.0000000000000916]
9. Dreher PC, Tong C, Ghiraldi E, Friedlander JI. Use of Google Trends to track online behavior and interest in kidney stone surgery. Urology 2018 Nov;121:74-78. [doi: 10.1016/j.urology.2018.05.040] [Medline: 30076945]
10. Ikpeze TC, Mesfin A. Interest in orthopedic surgery residency: a Google Trends analysis. J Surg Orthop Adv 2018;27(2):98-101. [Medline: 30084815]
11. Skopelja EN, Whipple EC, Richwine P. Reaching and teaching teens: adolescent health literacy and the internet. Journal of Consumer Health On the Internet 2008 Jun 17;12(2):105-118. [doi: 10.1080/15398280802121406]
12. Harmes KM, Blackwood RA, Burrows HL, Cooke JM, Harrison RV, Passamani PP. Otitis media: diagnosis and treatment. Am Fam Physician 2013 Oct 01;88(7):435-440 [FREE Full text] [Medline: 24134083]
13. Morley G. Otitis externa. Br Med J 1938 Feb 19;1(4024):373-377 [FREE Full text] [doi: 10.1136/bmj.1.4024.373] [Medline: 20781253]
14. Lieberthal AS, Carroll AE, Chonmaitree T, Ganiats TG, Hoberman A, Jackson MA, et al. The diagnosis and management of acute otitis media. Pediatrics 2013 Mar;131(3):e964-e999. [doi: 10.1542/peds.2012-3488] [Medline: 23439909]
15. Schaefer P, Baugh R. Acute otitis externa: an update. Am Fam Physician 2012 Dec 01;86(11):1055-1061 [FREE Full text] [Medline: 23198673]
16. Brkic FF, Besser G, Janik S, Gadenstaetter AJ, Parzefall T, Riss D, et al. Peaks in online inquiries into pharyngitis-related symptoms correspond with annual incidence rates. Eur Arch Otorhinolaryngol 2021 May;278(5):1653-1660 [FREE Full text] [doi: 10.1007/s00405-020-06362-4] [Medline: 32968893]
17. Knopke S, Böttcher A, Chadha P, Olze H, Bast F. [Seasonal differences of tympanogram and middle ear findings in children] German version. HNO 2017 Aug;65(8):651-656. [doi: 10.1007/s00106-016-0287-7] [Medline: 27904919]
18. Castagno LA, Lavinsky L. Otitis media in children: seasonal changes and socioeconomic level. International Journal of Pediatric Otorhinolaryngology 2002 Feb;62(2):129-134. [doi: 10.1016/s0165-5876(01)00607-3]
19. Gordon MA, Grunstein E, Burton WB. The effect of the season on otitis media with effusion resolution rates in the New York Metropolitan area. Int J Pediatr Otorhinolaryngol 2004 Feb;68(2):191-195. [doi: 10.1016/j.ijporl.2003.10.011] [Medline: 14725986]
20. Gharaghani M, Seifi Z, Zarei Mahmoudabadi A. Otomycosis in Iran: a review. Mycopathologia 2015 Jun;179(5-6):415-424. [doi: 10.1007/s11046-015-9864-7] [Medline: 25633436]
21. Centers for Disease ControlPrevention (CDC). Estimated burden of acute otitis externa--United States, 2003-2007. MMWR Morb Mortal Wkly Rep 2011 May 20;60(19):605-609 [FREE Full text] [Medline: 21597452]

22. Lu YX, Liang JQ, Gu QL, Yu XM, Yan X. [Study on the correlation between meteorological factors and acute otitis media in outpatients of children in Beijing]. Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi 2017 Oct 07;52(10):724-728. [doi: 10.3760/cma.j.issn.1673-0860.2017.10.002] [Medline: 29050087]
23. Crowson M, Schulz K, Tucci D. National utilization and forecasting of ototopical antibiotics: Medicaid data versus "Dr. Google". Otology Neurotology 2016;37(8):1049-1054. [doi: 10.1097/mao.0000000000001115]
24. FAQ about Google Trends data. Google. URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2021-01-07]
25. Adrian GB, Annette JD. Analysing Seasonal Health Data. Berlin, Heidelberg, Germany: Springer; 2009.
26. van Rossum G. Python tutorial. Centrum Wiskunde & Informatica Stichting.: Stichting Mathematisch Centrum; 1995. URL: https://ir.cwi.nl/pub/5007 [accessed 2021-06-01]
27. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020 Sep;585(7825):357-362 [FREE Full text] [doi: 10.1038/s41586-020-2649-2] [Medline: 32939066]
28. The Pandas development team. Pandas. URL: https://pandas.pydata.org/about/team.html [accessed 2021-04-27]
29. Hunter JD. Matplotlib: a 2D graphics environment. Comput. Sci. Eng 2007 May;9(3):90-95. [doi: 10.1109/mcse.2007.55]
30. Jordahl K. GeoPandas: Python tools for geographic data. GeoPandas 0.9.0. URL: https://geopandas.org/ [accessed 2021-04-27]
31. Patthi B. Global search trends of oral problems using Google Trends from 2004 to 2016: an exploratory analysis. JCDR 2017 Sep 1:C12-C16. [doi: 10.7860/jcdr/2017/26658.10564]
32. Shen JK, Every J, Morrison SD, Massenburg BB, Egbert MA, Susarla SM. Global interest in oral and maxillofacial surgery: analysis of Google Trends data. J Oral Maxillofac Surg 2020 Sep;78(9):1484-1491. [doi: 10.1016/j.joms.2020.05.017] [Medline: 32554065]
33. Harorli OT, Harorli H. Evaluation of internet search trends of some common oral problems, 2004 to 2014. Community Dent Health 2014 Sep;31(3):188-192. [Medline: 25300156]
34. Samaras L, García-Barriocanal E, Sicilia M. Syndromic surveillance models using web data: the case of influenza in Greece and Italy using Google Trends. JMIR Public Health Surveill 2017 Nov 20;3(4):e90 [FREE Full text] [doi: 10.2196/publichealth.8015] [Medline: 29158208]
35. Zhang Y, Bambrick H, Mengersen K, Tong S, Hu W. Using Google Trends and ambient temperature to predict seasonal influenza outbreaks. Environ Int 2018 Aug;117:284-291. [doi: 10.1016/j.envint.2018.05.016] [Medline: 29778013]
36. Schnoell J, Besser G, Jank BJ, Bartosik TJ, Parzefall T, Riss D, et al. The association between COVID-19 cases and deaths and web-based public inquiries. Infect Dis (Lond) 2021 Mar;53(3):176-183. [doi: 10.1080/23744235.2020.1856406] [Medline: 33287607]
37. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Jul;10(7):839-847 [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]
38. Venkatesh U, Gandhi PA. Prediction of COVID-19 outbreaks using Google Trends in India: a retrospective analysis. Healthc Inform Res 2020 Jul;26(3):175-184 [FREE Full text] [doi: 10.4258/hir.2020.26.3.175] [Medline: 32819035]
39. Ayyoubzadeh S, Ayyoubzadeh S, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: data mining and deep learning pilot study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
40. Marchetti F, Ronfani L, Nibali SC, Tamburlini G, Italian Study Group on Acute Otitis Media. Delayed prescription may reduce the use of antibiotics for acute otitis media: a prospective observational study in primary care. Arch Pediatr Adolesc Med 2005 Jul;159(7):679-684. [doi: 10.1001/archpedi.159.7.679] [Medline: 15997003]
41. Jacobs MR, Dagan R, Appelbaum PC, Burch DJ. Prevalence of antimicrobial-resistant pathogens in middle ear fluid: multinational study of 917 children with acute otitis media. Antimicrob Agents Chemother 1998 Mar;42(3):589-595 [FREE Full text] [doi: 10.1128/AAC.42.3.589] [Medline: 9517937]
42. Arrieta A, Singh J. Management of recurrent and persistent acute otitis media: new options with familiar antibiotics. Pediatr Infect Dis J 2004 Feb;23(2 Suppl):S115-S124. [doi: 10.1097/01.inf.0000112525.88779.8b] [Medline: 14770074]
43. Kaushik V, Malik T, Saeed S. Interventions for acute otitis externa. Cochrane Database Syst Rev 2010;1. [doi: 10.1002/14651858.cd004740.pub2]
44. Roland PS, Stroman DW. Microbiology of acute otitis externa. Laryngoscope 2002 Jul;112(7 Pt 1):1166-1177. [doi: 10.1097/00005537-200207000-00005] [Medline: 12169893]
45. Laulajainen-Hongisto A, Aarnisalo AA, Jero J. Differentiating acute otitis media and acute mastoiditis in hospitalized children. Curr Allergy Asthma Rep 2016 Oct;16(10):72. [doi: 10.1007/s11882-016-0654-1] [Medline: 27613655]
46. van den Aardweg MTA, Rovers M, de Ru JA, Albers F, Schilder A. A systematic review of diagnostic criteria for acute mastoiditis in children. Otol Neurotol 2008 Sep;29(6):751-757. [doi: 10.1097/MAO.0b013e31817f736b] [Medline: 18617870]
47. Stalfors J, Enoksson F, Hermansson A, Hultcrantz M, Robinson, Stenfeldt K, et al. National assessment of validity of coding of acute mastoiditis: a standardised reassessment of 1966 records. Clin Otolaryngol 2013 Apr;38(2):130-135. [doi: 10.1111/coa.12108] [Medline: 23577881]
48. Psarommatis IM, Voudouris C, Douros K, Giannakopoulos P, Bairamis T, Carabinos C. Algorithmic management of pediatric acute mastoiditis. Int J Pediatr Otorhinolaryngol 2012 Jun;76(6):791-796. [doi: 10.1016/j.ijporl.2012.02.042] [Medline: 22405736]

49. Quesnel S, Nguyen M, Pierrot S, Contencin P, Manach Y, Couloigner V. Acute mastoiditis in children: a retrospective study of 188 patients. Int J Pediatr Otorhinolaryngol 2010 Dec;74(12):1388-1392. [doi: 10.1016/j.ijporl.2010.09.013] [Medline: 20971514]
50. Amaro CE, Espiney R, Radu L, Guerreiro F. Malignant (necrotizing) externa otitis: the experience of a single hyperbaric centre. Eur Arch Otorhinolaryngol 2019 Jul;276(7):1881-1887. [doi: 10.1007/s00405-019-05396-7] [Medline: 31165255]
51. Babiatzki A, Sadé J. Malignant external otitis. J Laryngol Otol 1987 Mar;101(3):205-210. [doi: 10.1017/s0022215100101549] [Medline: 3106545]
52. Pichichero ME. Recurrent and persistent otitis media. Pediatr Infect Dis J 2000 Sep;19(9):911-916. [doi: 10.1097/00006454-200009000-00034] [Medline: 11001126]
53. Bhattacharyya N, Kepnes LJ. Initial impact of the acute otitis externa clinical practice guideline on clinical care. Otolaryngol Head Neck Surg 2011 Sep;145(3):414-417. [doi: 10.1177/0194599811406797] [Medline: 21531870]
54. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Jul;10(7):839-847 [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]
55. Telfer S, Woodburn J. Let me Google that for you: a time series analysis of seasonality in internet search trends for terms related to foot and ankle pain. J Foot Ankle Res 2015;8:27 [FREE Full text] [doi: 10.1186/s13047-015-0074-9] [Medline: 26146521]

Abbreviations

GT: Google Trends
ICC: intraclass correlation coefficient
OE: otitis externa
OM: otitis media
RSV: relative search volume

Edited by R Kukafka; submitted 01.03.21; peer-reviewed by T Cruvinel, S Kardes, JK Kumar, C Hudak; comments to author 13.04.21; revised version received 27.04.21; accepted 06.05.21; published 20.06.21.

Please cite as:
Brkic FF, Besser G, Schally M, Schmid EM, Parzefall T, Riss D, Liu DT
Biannual Differences in Interest Peaks for Web Inquiries Into Ear Pain and Ear Drops: Infodemiology Study
J Med Internet Res 2021;23(6):e28328
URL: https://www.jmir.org/2021/6/e28328/
doi:10.2196/28328
PMID:34185016

Original Paper

Mobile Health in Chronic Disease Management and Patient Empowerment: Exploratory Qualitative Investigation Into Patient-Physician Consultations

Kathrine Stampe[1], PhD; Sharon Kishik[1], BA; Sune Dueholm Müller[1], PhD

Department of Management, School of Business and Social Sciences, Aarhus University, Aarhus V, Denmark

Corresponding Author:
Sune Dueholm Müller, PhD
Department of Management
School of Business and Social Sciences
Aarhus University
Fuglesangs Allé 4
Aarhus V, DK-8210
Denmark
Phone: 45 61776176
Email: sdm@mgmt.au.dk

Abstract

Background: Chronic diseases often present severe consequences for those affected. The management and treatment of chronic diseases largely depend on patients' lifestyle choices and how they cope with the disease in their everyday lives. Accordingly, the ability of patients to self-manage diseases is a highly relevant topic. In relation to self-management, studies refer to patient empowerment as strengthening patients' voices and enabling them to assert control over their health and treatment. Mobile health (mHealth) provides cost-efficient means to support self-management and foster empowerment.

Objective: There is a scarcity of research investigating how mHealth affects patient empowerment during patient-physician consultations. The objective of this study is to address this knowledge gap by investigating how mHealth affects consultations and patient empowerment.

Methods: We relied on data from an ethnographic field study of 6 children and adolescents diagnosed with juvenile idiopathic arthritis. We analyzed 6 patient-physician consultations and drew on Michel Foucault's concepts of power and power technology.

Results: Our results suggest that the use of mHealth constitutes practices that structure the consultations around deviations and noncompliant patient behavior. Our analysis shows how mHealth is used to discipline patients and *correct* their behavior. We argue that the use of mHealth during consultations may unintentionally lead to relevant aspects of patients' lives related to the disease being ignored; thus, inadvertently, patients' voices may be silenced.

Conclusions: Our results show that concrete uses of mHealth may conflict with extant literature on empowerment, which emphasizes the importance of strengthening the patients' voices and enabling patients to take more control of their health and treatment. We contribute to the state-of-the-art knowledge by showing that the use of mHealth may have unintended consequences that do not lead to empowerment. Our analysis underscores the need for further research to investigate how mHealth impacts patient empowerment during consultations.

(J Med Internet Res 2021;23(6):e26991) doi:10.2196/26991

KEYWORDS

compliance; empowerment; mHealth; patient-physician consultation; power

Introduction

Background

Chronic diseases often have severe consequences for those affected. In addition to the reduced quality of life, chronic diseases may lead to health emergencies, serious complications, and even death [1]. The management of chronic diseases is among the many costly challenges faced by the health care sector [2]. Each year, chronic diseases account for 71% of all deaths globally, prompting the World Health Organization to call for immediate action [3]. Furthermore, chronic diseases contribute to inequality, as they disproportionally affect socially disadvantaged people and impede poverty reduction initiatives [3]. Hence, improving the health of patients with chronic diseases not only improves their quality of life but also eases economic burdens at a societal level. The management and treatment of chronic diseases depend on their impact on patients' everyday lives [4,5]. Improving patients' abilities to self-manage their health and make informed decisions is recognized as the key to manage chronic disease at a societal level. This is reflected in numerous national health care strategies that focus on personalized medicine and self-management of disease [6,7]. Self-management refers to patients' active engagement in the responsibility for their health [1]. In relation to self-management, studies refer to patient empowerment as enabling patients to more meaningfully engage with and assert control over their health and treatment of diseases [8-10]. To support self-management and foster empowerment, mobile health (mHealth) is a cost-efficient method [11].

mHealth is the delivery of health services and information over a mobile or wireless platform [12]. The literature on mHealth in the chronic disease self-management describes various technologies that have been introduced in health care to support self-management and foster patient empowerment [13]. Among these are mobile apps that allow for the collection, management, and sharing of patient-reported outcome (PRO) data. PRO data are direct responses from the patients' perspectives regarding their health condition, without interpretation from a provider or caregiver [14,15]. Traditionally, PRO data have been collected during physical consultations. However, consultations are often months apart, resulting in recall bias and unreliable data. mHealth may offer a solution by allowing continuous patient input and consequently more reliable PRO data [16]. During patient-physician consultations, PRO data support physicians in evaluating the health status of patients. However, despite studies of how PRO data and mHealth support patient empowerment [17-19], there is a scarcity of research on how mHealth and other technologies are used during patient-physician consultations, and extant studies often ignore the fact that empowerment depends on the patient-physician relationship [8,20]. Consequently, there is limited knowledge of how and to what extent the use of mHealth during consultations affects empowerment. This study addresses this knowledge gap through an exploratory investigation of consultations between children and adolescents diagnosed with juvenile idiopathic arthritis (JIA) and their physician. During these consultations, an app named *How-R-you* supports decision making and patient-physician communication. For analysis purposes, we draw on a theoretical framework based on Michel Foucault's concepts of power and power technology. Using this framework, we show how the exercise of power is embedded in the different uses of mHealth that constitute specific practices during consultations. In our investigation, we were guided by the following research question:

How is mHealth technology used during patient-physician consultations and to what extent does it support patient empowerment?

The remainder of this paper is organized as follows. First, we account for state-of-the-art knowledge of mHealth in chronic disease self-management and the role of patient empowerment. Second, we present the theoretical framework. Third, we introduce the How-R-you case study and account for our approach to data collection and analysis. Finally, we present our results and discuss their contributions and implications.

Background Literature

The literature on mHealth in chronic disease self-management covers various technologies and platforms. Technologies include wearables [21], social media platforms [22], and mobile apps [23]. Technologies are studied mainly from one of two perspectives: health care professionals and patients.

From the perspective of a health care professional, mHealth technology provides health care services over geographical distances and supports patient self-management. mHealth may be used to transmit patient data to health care professionals [24] and provide health services in remote areas [25]. Paré et al [24] reported positive effects such as a decreased need for emergency care. Studies also point to greater empowerment, but results regarding clinical outcomes and costs are inconclusive [26].

From a patient perspective, mHealth technology provides many benefits because of the widespread use of mobile phones and the possibility of integrating data from various wearable devices [27]. Patients can monitor symptoms and health issues, which may help them achieve their health objectives [28]. Moreover, various technologies serve as external memories that help patients in remembering health details [29]. This, in turn, supports patients in self-managing their health and coping with diseases [2]. Furthermore, mHealth enables information sharing between patients and health care professionals [19]. As a result, patients increase their knowledge, which helps them to self-manage chronic diseases [19,23]. However, studies have reported problems such as constantly reminding patients that they are chronically ill [30]. Moreover, patients often avoid using technologies because their values and needs are ignored in the design of mHealth [1]. The solution is argued to be a patient-centered technology that strengthens the patients' voices [1]. In the literature, this is often referred to as a technology that supports patient empowerment.

The Role of Patient Empowerment

Empowerment implies that patients can fulfill their needs, assert control of their treatment, and independently—or at the very least as an equal partner—decide on appropriate behavior and treatment [10,31,32]. In a widely recognized perspective on empowerment, patients' subjective experiences of the disease are central to follow-up and treatment. By strengthening patients' voices, empowerment is an engagement of patients that goes beyond mere compliance [33]. However, in much of the literature, empowerment is also conceptualized as the measurable result of an intervention, and the role of mHealth technology in empowering patients is thus paradoxically referred to, discussed, and measured in terms of compliance with established norms and treatment [34]. For instance, Lasorsa et al [2] designed an mHealth technology to support patients' self-management and everyday decision making. The solution is said to incorporate empowering features and a means to ensure compliance with medical treatment. Similarly, Fioravanti et al

[35] studied an mHealth strategy aiming to improve medical compliance through empowerment. This strategy promotes lifestyle changes and improves compliance. With such studies, the literature articulates a paradoxical conceptualization of empowerment. Patients gain control in the sense that they improve their ability to comply with the physicians' medical advice. Following this logic, empowerment is less about patients' everyday lives and their subjective experiences with the disease than their compliance with treatment plans. According to this conceptualization, empowerment is disease-rather than patient-centered, and technologies continue to privilege a clinical perspective [33].

Empowerment must be patient-centered if the patients' voices are to be truly strengthened. Dadgar and Joshi [1] argue that mHealth must cater to patients' values to affect empowerment. Butterworth et al [36] combined patient and clinical perspectives to create a technology-enabled education system that, on one hand, serves patients' needs by improving their ability to make informed decisions and, on the other hand, supports health care professionals by teaching patients to care for themselves and reminding them of appointments. Storni [33] proposed a patient-centered mHealth journaling tool that considers patients' individual needs and enables self-monitoring personalization. Technologies offering privilege to the health care professional's perspective assume that health care outcomes are always clinically measurable and necessitate compliance: "Such a strategy may verge on the paradox of 'empowering' patients to better silence their voices, but poorly supports them with the practicalities and complexities of dealing with the disease" [33]. Patients experience diseases subjectively, and thus should not be reduced to generic medical conditions [33]. Otherwise, patients' insights and knowledge are neglected at the risk of ignoring relevant aspects of their lives with the diseases. This is particularly important as patients' lifestyle choices and their everyday lives are key to managing chronic diseases [5]; hence, the emphasis on empowerment as strengthening patients' voices and enabling patients to take control of their health and treatment.

Although the referenced studies constitute important contribution to the literature on digitally enabled chronic disease self-management, the extant literature does not investigate the role of mHealth technology during patient-physician consultations, and studies thus tend to ignore that empowerment is dependent on the patient-physician relationship [20]. Consequently, the state-of-the-art knowledge of how to use mHealth technology and PRO data during consultations, and whether they promote patient empowerment, is insufficient. Use of mHealth that reduces patients to generic medical conditions is not conducive to empowerment and may result in patient-physician miscommunication [37]. Therefore, research is needed to investigate how mHealth is used and how it affects empowerment. This paper addresses this knowledge gap and analyzes the use of mHealth technology during patient-physician consultations.

Methods

Theoretical Framework

We analyzed the use of mHealth through a Foucauldian perspective. Foucault's concept of power as embedded in institutional practices and exercised over free subjects is particularly useful while trying to understand how power is exercised in modern state institutions (eg, hospitals) and investigating the aforementioned tensions between compliance and empowerment [38,39].

Foucault's understanding of power stands in stark contrast with the conceptualization of power as an oppressive force [40]. Instead, power is productive as it creates subjects with specific characteristics and attributes. Moreover, power cannot be possessed, and its locus cannot be pinpointed with reference to a person in a leading position. Power *is* only insofar as it is exercised, meaning it emerges only in relationships, "...as a mode of action upon the actions of others" [40]. Thus, it presupposes the ability of subjects to act freely. Therefore, the study of power must be conducted as an analysis of the relational practices through which some are directed to adjust their behavior according to certain norms [40]. An example of such an analysis is found in "Discipline and Punish" [41]. Here, Foucault analyzes discipline as a form of power that gains ground in the 19th century, and which—as we will show—is being continuously refined today. His analysis takes as its objects the various practices in prisons, through which the inmates are governed to discipline and adjust their behavior. Discipline is a power that rectifies norm-deviant behavior by initiating normalizing processes. By the same act, it creates productive yet docile subjects [41].

For the purposes of our analysis of mHealth use during patient-physician consultations, Foucault's concept of power technology provides a potent analytical tool. Power technology refers to something that is mobilized in specific contexts to govern the conduct of others. Power technology structures and organizes contexts in specific ways. It may render some things visible, whereas others are left invisible or hidden [42]. Nothing is a power technology in and of itself. It comes to constitute a power technology, only when, in a certain context, it is invested with a specific rationality and used to govern the conduct of others [43]. This does not mean, however, that mHealth *is* inherently a power technology. Rather, construing it as a power technology enables us to analyze how mHealth organizes consultations and is used to govern behavior. Only as far as it organizes consultations in particular ways, mHealth is constituted as a power technology.

In this study, we drew on Foucault's concepts of power and power technology to analyze the different uses of the How-R-you app that constitute specific practices during consultations. Thus, although other frameworks conceive power as something that can be possessed and used for different purposes, the Foucauldian perspective allows us to analyze how power emerges subtly through health care practices. We may thereby gain new insights into how technology, which is seemingly empowering patients, also governs their behavior in specific ways.

mHealth as a Power Technology

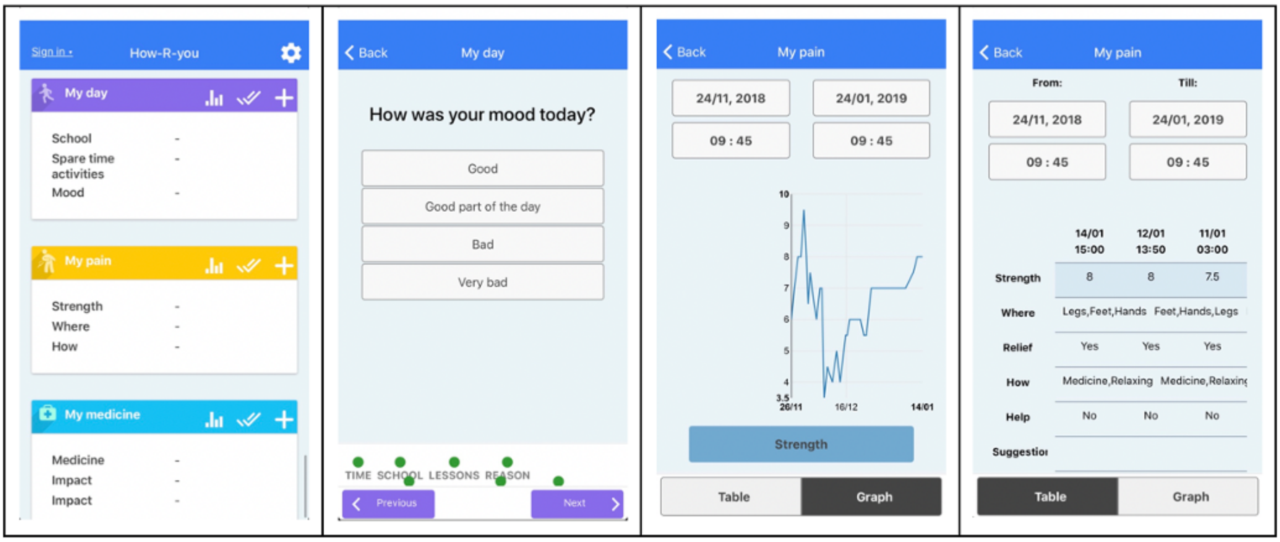
The propagation of mHealth and other technologies, meant to engage and empower patients, must be viewed in light of the historical development within health care, with an increased focus on patient-centered care. It has been argued that we have seen a crisis in traditional management and authority, and that physicians must increasingly refrain from exercising authoritative, paternalistic power. Instead, the patient-physician relationship must be based on dialogue [42]. Especially with the approach often taken to chronic diseases, where patients' lifestyle choices are in focus, the physician must incorporate the patient's perspective, on account of the premise that any real lifestyle changes come from within the patient [42,44]. This creates challenges in the management of patients with chronic diseases: health care must be provided without commanding or ordering patients.

In this paper, we argue that mHealth is sometimes used as a solution to this challenge. mHealth technologies invoke a seemingly unoppressive power because they provide patients the right to speak and be heard, which is perceived as liberating and empowering [42]. They give patients a voice; yet, they are a means to ensure compliance. Specific practices during consultations highlight this point. Thus, we argue that mHealth does not simply liberate patients from being subjugated to power; rather, it reinstalls patients in new power relations, in which their seemingly empowered voice is used to govern them in specific directions. Thus, mHealth governs by ensuring that the patients are, in Rose's terms, "bound into the language of expertise at the very moment they are assured of their freedom and autonomy" [45]. In other words, the patients are paradoxically governed exactly through the freedom that mHealth seems to provide them. In the following analysis, we argue that How-R-you is constituted as a power technology. Although it seems liberating, it also organizes consultations to ensure patient compliance according to medical recommendations. Thus, it is invested with a governmental rationale.

How-R-you

How-R-you was chosen for this study because it was developed with a focus on empowerment and strengthening of patients' voices. The app was developed by Business Academy Aarhus with input from patients, health care professionals from Aarhus University Hospital, and researchers from Aarhus University. The app was commissioned by physicians to gain a better understanding of patients' everyday lives with chronic diseases and the symptoms they experience. How-R-you allows patients to continuously report on their health and monitor their well-being using PRO data. Specifically, the app enables 4 of 7 self-management activities [46] in patients' everyday life (drug management, communication with health care professionals, social support, and symptom management) by allowing patients to track drug use and monitor symptoms, and by supporting communication with both health care professionals and other stakeholders, especially parents [47]. How-R-you is organized into modules that contain health questions. The modules are "My medicine," "My pain," "My day," and "My night" (Figure 1).

Figure 1. How-R-you modules, question, graph, and table examples.



How was your mood today?

	14/01 15:00	12/01 13:50	11/01 03:00
Strength	8	8	7.5
Where	Legs,Feet,Hands	Feet,Hands,Legs	
Relief	Yes	Yes	Yes
How	Medicine,Relaxing	Medicine,Relaxing	
Help	No	No	No
Suggestion			

Modules and questions were developed and revised iteratively, and both patients' and physicians' needs were included in the design. The modules are configurable, which enables personalization according to the specific preferences of patients and physicians. Furthermore, How-R-you allows patients and physicians to gain an overview of historical data by converting them into graphs and tables.

Data Collection

We relied on data from a 2-year ethnographic field study of the impact of digital technology use in everyday life with JIA. The ethnographic approach helped us understand real-world problems empirically before seeking to explain them theoretically [48-50]. The fieldwork yielded both overall and individual accounts of 6 patients' everyday lives with JIA and their visits to the hospital. These accounts constituted thick descriptions [51] of patients' information needs and the extent to which How-R-you as an mHealth technology created value for patients [47]. The fieldwork comprised participant observations and interviews. Following Spradley [52], observations are always participatory, but to varying degrees. Observations were descriptive at the outset of the study to gain insights into the sociotechnical aspects of technology use. The researchers participated passively at the outset. Subsequently, observations became more focused, and participation became more involved in the sense of taking part in patients' daily activities. Thus, additional data were collected to gain a deeper understanding of selective aspects, that is, how How-R-you is used during consultations. The data were collected by the second author (KS), with support from the third author (SDM) who participated in the hospital. For this study, we carefully selected 6 specific observations of consultations as the empirical basis for our analysis. All observations were carefully planned and subsequently discussed to ensure reliable and valid interpretations.

The 6 cases comprised 1 consultation with each of the 6 patients included in the ethnographic study [47]. Patients were selected based on four criteria. First, all patients had the same disease (ie, JIA). Second, they were not diagnosed with other diseases (comorbidities). Third, both female and male patients were included. Finally, 1 female and 1 male patient were selected to represent each of the three predetermined maturity groups (7-10, 11-13, and 14-17 years old) that were a part of the study. The age groups were based on literature describing the medical assessments of JIA children and adolescents' cognitive and physiological development [53]. Children can assess their pain using a visual analogue scale, as in the case of How-R-you, from age 6 [54,55]. Furthermore, patients aged 18 years and above were treated as adults at Aarhus University Hospital. This means that our target group was 6-17 years old. In addition, the literature divides JIA children into 2 age groups; however, the cutoff age varies [53] between 12 [56] and 13 years [57] (Table 1).

Table 1. Overview of patients.

Patient	Sex	Age (years)
Nina	Female	7
Robert	Male	10
Rachel	Female	12
Oliver	Male	12
John	Male	14
Sophie	Female	16

We specifically selected these 6 consultations because they were held during the final stages of the fieldwork. During this stage of the fieldwork, both patients and the physician had become relatively familiar with How-R-you, and, therefore, technical questions and issues were less frequent. Moreover, at this selective stage of the fieldwork, we relied on extensive experience after 2 years of descriptive and focused interviews and observations of the patients' lives. During this time, the app underwent iterative development because of feedback from patients and physicians. Consequently, at this stage, we focused our observations specifically on the actual use of How-R-you during consultations [47]. Thus, these 6 consultations provide an ideal basis for our investigation into how How-R-you is used during consultations and how it affects empowerment.

All consultations were performed by the same physician to limit in- and cross-case variations. Although all the children's parents were present at the consultations, their involvement varied across the age groups. Parents of young children were highly involved, whereas parents of adolescents were not involved [47]. The duration of consultations varied between 20 and 30 minutes depending on the treatment needs and issues raised during conversation with the physician. The consultations were documented through participant observation [52], audio-recorded, and transcribed. Transcriptions were performed by the first (SK) and second (KS) authors. The consultations were discussed in detail to ensure that all aspects were included in the analysis. The observations were carefully planned with both the physician and patient. In accordance with the nature of passive participation, we interfered as little as possible in the consultations and only interacted with participants when encouraged by patients or the physician. To minimize the disruptive effects of our presence, we took field notes with pen and paper from the corner of the consultation room. Field notes were available to the participants.

Ethics

All patients, parents (due to the patients' ages), and the physician gave informed consent to all parts of the fieldwork. The physician collaborated with the research team throughout the fieldwork and participated in discussions about ethics. The patients and their parents were invited to a private informal meeting with one of the authors, who explained the details of the research projects. At least one week later, the author called patients and parents, who had expressed an interest in joining the study at the meeting. This gave patients and parents an opportunity to reconsider their interests and ask questions about the research project. Consent was given at the end of this conversation and more specific permissions, such as audio recording and taking pictures, were obtained throughout the entire study. The study complies with the ethical principles for medical research involving human subjects and with the American Anthropological Association's ethical principles for ethnographic fieldwork [58]. The study did not require approval from an ethical review committee according to the Danish National Committee on Biomedical Research Ethics.

Data Analysis

To address the research question, we draw on Foucault's concepts of power and power technology in an embedded case study of patient-physician consultations. This means that we seek to understand how How-R-you is used during consultations, what kind of practices it constitutes, and the rationale that drives the organization of these practices.

In our analysis, we are not concerned with the physician's intentions and motivations behind the use of the app, but only with how it is used and what it does to the consultation and the patient-physician relationship. Our interest lies not in the participants but in how the practices themselves lay the foundation for the exercise of power by organizing the consultation and inducing a certain behavior. Thus, How-R-you is not used by someone *in* power to assert power. Rather, power emerges as the way in which How-R-you organizes the relationship between the patient and the physician (Textbox 1).

Textbox 1. Data coding.

Identification of Deviation

- Pain
- Disturbed sleep
- Bad mood
- Physical inactivity

Structuring of Conversation

- Specifics about the deviation
- Patterns between disease indicators
- Ways of handling the deviation

Discipline

- Education
- Commending
- Questioning

Consultations were analyzed through an iterative process. By observing the consultations and listening to the recordings numerous times, we gained a preliminary understanding of the practices. Next, we transcribed and coded the data using the qualitative data analysis software NVivo 12 (QSR International). Top-level codes were identified based on our preliminary understanding of practices and our theoretical framework. This means that we sought to understand how the app is used organizes certain practices in specific ways and how it is invested with a specific rationale. Thus, we asked the following questions: How is the app used in specific practices? What registrations become visible to the physician and what are left unmentioned? How are the registrations used in conversation? Which topics become important during the conversation, and which are rendered irrelevant?

All authors were involved in data coding and analysis. The first (SK) and second (KS) authors provided deep insights and domain knowledge from participant observations and interviews. The third author (SDM) contributed to the in-depth text analysis of communication content, check coding, and challenging the interpretations of other authors. Through this process, we meticulously went through the transcriptions and coded the data according to the top-level codes. This led to the identification of sublevel codes for each top-level code. The process provided an in-depth understanding of the consultations and practices described in the following analysis.

Results

Overview

Our analysis reveals three interrelated practices around the use of How-R-you: ordering the patients' words; identifying deviations and structuring of patients' speech; and disciplining of patients. In the following, we argue that these practices collectively constitute How-R-you as a power technology. We devote each of the following three subsections to each specific practice. Finally, we conclude this chapter by summarizing the results.

Ordering the Patients' Words

How-R-you allows patients to register and monitor their health data. These data were examined by the physician and included in the health assessments of the patients during consultations. Therefore, How-R-you is believed to strengthen patients' voices by giving them the ability to provide personal insights. Before the introduction of How-R-you, patients arrived at consultations without a record of their health and well-being since their last consultations. In fact, patients and parents (who support their children) rarely take notes on their health from day to day. The result is recall bias when asked to account for their well-being since the last consultation. In How-R-you, patients make continuous health registrations by answering specific questions in the app. Thus, How-R-you constructs and builds a record of PRO data. During consultations, the data were presented to the physician as the patients' own words. However, in How-R-you, the registrations are ordered in certain ways, for example, in graphs that show how the patient felt each day since his or her last consultation:

> *Then we can use the registrations you have made (...) we just need to construct the graphs, because, you know, that's what's important—how your day has been.* [Physician]

The technology itself becomes an actor with a voice of its own. It simplifies patient experiences and confines registrations to specific questions. For instance, moods and feelings, which may be difficult to explain, are registered using smiley and slider scales (Figure 1). These registrations are presented in tables and graphs that allow physicians to examine the data. As a result of this ordering of the patients' words, physicians can quickly sum up otherwise complex experiences. For example, in the following quotation, a patient's mood and physical activity since the last consultation are summarized as follows:

> *Well, your mood seems to have been okay despite some tiresome nights. And your level of activity has been normal.* [Physician]

Thus, How-R-you is not a neutral instrument. It allows patients to continuously register health observations by answering specific questions, and it presents these registrations in a particular manner. The way How-R-you is used constitutes a practice that orders patients' words but presents them as their own voice. This allows for a seemingly empowering process through which patients' insights are incorporated during consultations. However, the ordering of the patients' words only enables very specific observations of the patients' lives with the disease. In the following section, we expand on this last point.

Identification of Deviations and Structuring of Patients' Speech

The ordering of patients' words exerts a structuring effect on conversations and patients' speech during consultations. During the physician's examination of data, it is evident that the physician scans the app modules for deviations, such as pain, disturbed sleep, mood swings, or physical inactivity. In other words, How-R-you renders deviations visible to physicians. How-R-you structures the conversation around the observed deviations. It does not reduce conversation as it expands it around very specific topics. With reference to their registrations in How-R-you, the patients were encouraged to speak and describe the deviations. Their speech is therefore structured by limiting the conversation to (1) specifics regarding the deviation, (2) patterns across disease indicators, and (3) how they *handle* the deviation:

> *What should I look at...your pain? Do you ever feel pain? Yes, you do actually—pain in the legs.* [Physician]

The above quote is from Nina's consultation, where the physician browsed the "My pain" module. The physician observed that Nina registered pain in her legs. The conversation then unfolds around this deviation. The physician probed the specifics of the pain, and Nina's mother described when the pain usually occurs and what they did to relieve it. In other words, it is about how they *handle* the deviation. This conversation helped the physician ascertain the cause of the pain.

In Robert's case, the physician's data analysis also revealed pain, and the conversation again focused on this topic. Neither Robert nor his father could remember the specific incidents where Robert experienced pain. However, Robert's father was uncertain whether the pain is attributable to arthritis or because Robert engaged in new and intense physical activities. This indicates that pain is not only related to arthritis:

> *...when it is something [pain] with the hand, well, that sounds like it's because you do something that you regularly don't, right?* [Physician]

During Rachel's consultation, the physician observed that Rachel had registered mood swings. However, Rachel attributed the cause to "...something at school," which seemed to indicate the physician that her mood was not relevant to the consultation. The physician continued by looking at other data in How-R-you and suggested that mood swings are connected to another disease indicator, namely physical inactivity. Similarly, Sophie attributed her mood swings to the lack of energy caused by a combination of pain and stress at work. In this case, the physician ascertained through How-R-you that there was no correlation with other disease indicators. The physician then told her that to cope with these mood swings, she will have to figure out what she can and cannot do:

> *That's part of what's going to be your challenge—to figure out the balancing act in what you are able to do and what you cannot do.* [Physician]

In the examples above, the structuring of conversations around deviations and the patients' own accounts led the physician to draw conclusions. The examples show that deviations are well-handled (Nina), do not relate directly to the arthritis (Robert and Sophie), or seem not relevant to the consultation (Rachel).

Notably, although the patients' perspectives have been included in the design of How-R-you to foster empowerment and strengthen the patients' voices, the app seems to structure the patient-physician consultations in a manner that seems to provoke somewhat opposite effects. One may argue that during the consultations, How-R-you renders certain relevant aspects of patients' lives irrelevant. These are aspects that the patients themselves bring up or have registered in the app (eg, what happens at school), indicating that they are relevant to their lives and, thus, to the self-management of their disease. However, the way How-R-you organizes registrations, for example, in graphs and tables, seems to leave these aspects in the blind spot of the physician's gaze, which is focused on deviations and patterns across disease indicators. Data are shared, but patients' voices are only selectively included as a basis for disease management and treatment during consultations. This is because How-R-you only enables very specific observations of patients' lives with the disease. As our analysis suggests, these observations leave relevant aspects unnoticed.

In other cases, patients' speech becomes a reference point for the exercise of power and identified deviations become targets for corrective, disciplinary power. We provide examples in the following section.

Disciplining of Patients

Our argument in the following is that by facilitating, ordering, and structuring patients' words and speech, How-R-you allows for the exercise of disciplinary power over patients with reference to their own spoken truths. The physician avoids referring to her superior medical knowledge and position. Instead, by referring to patients' own registrations, disciplining is conducted without the exercise of oppressive power. Specifically, disciplining takes the form of education, commending "right" (read in a medical sense) behavior and questioning patients.

During John's consultation, the physician browsed the How-R-you registrations and observed a deviation regarding his sleep patterns:

> *This is your sleep. So, disturbed [sleep], then a little steady and then very [disturbed].* [Physician]

John was asked to account for details around his sleep and he responded by describing his tiredness. This deviation, observed through the app, and the patient's description form the background for the physician's comment:

> *...it's still two days, where your sleep was very disturbed, so that's something, where you should keep in mind that the best thing that sleep knows is consistency. So, you go to bed at the same time, you wake up at the same time, you don't change it too much in the weekends...* [Physician]

With reference to the deviation, educating the patient about how to improve his sleep is legitimized. In other words, education is necessary because "it's still two days, where your sleep was very disturbed."

The physician continues by describing optimal bedroom temperatures and warnings against the effects of technology use before sleep:

> *This means that all kinds of tablet, tv, computer, and all that stuff should preferably not be switched on [...] So I've just asked you to use an app [How-R-you], and now I'm asking you to really not use it too much before bedtime, or at least that's something to think about.* [Physician]

Thus, education and suggested behavioral changes are used to *correct* the deviation. Both John and his mother engaged in conversations with the physician and showed signs of compliance. John mentioned that he switches his iPhone lighting to *night shift*, which changes the display to warmer colors. His mother emphasized that they imposed stricter bedtime rules regarding when to turn off video games. In response, the physician replied: "Right, it's those video games." Thus, John and his mother show compliance by explaining how they are going to *correct* the deviant behavior. A part of disciplining the patient is to make him or her recognize the need for behavioral change. This involves a conversation about the deviation during which compliant behavior is constructed as *right* and noncompliant behavior (eg, playing games before bedtime) is articulated as *wrong*.

During Rachel's consultation, the topic of physical inactivity emerged because she had registered periods of pain. The physician noted that the number of days of pain correlated with days where she was less active. This topic was revisited later during consultation. Although the physician did not reference the observed deviation in the subsequent disciplining, there was a correspondence between the observed deviation (physical inactivity) in How-R-you and the discipline. Thus, the registrations in How-R-you reveal that pain may be caused by physical inactivity (the deviation). In this sense, the registrations form the background for the following conversation:

> *Yesterday, I walked 30,000 [steps], because I rode [my horse] twice, then I had physical education and rode my bike and stuff.* [Rachel]
>
> *Well-done [...] I thought you'd be at around 8,000-10,000 [steps] most of the times, unless you're one of those, who play Fortnite, right.* [Physician]
>
> *Right.* [Rachel]
>
> *Then it's at a zero.* [Physician]

Rachel's response revealed compliant behavior (physical activity). This behavior was encouraged by the physician who commended her while at the same time alluded to noncompliant (ie, *wrong*) behavior, namely inactivity and too much gaming. Here, discipline takes the form of commending compliant behavior and constructing noncompliant behavior as *wrong*.

During Oliver's consultation, the physician observed that Oliver had registered periods of pain. When asked about the specifics, he revealed deviant behavior in the form of physical inactivity:

> *So, you nevertheless had one day in January, where you felt pain in your leg.* [Physician]
>
> *Yes, so I had a 7.5 (on the pain intensity scale) [...]* [Oliver]
>
> *Was it related to football, or was it just kind of sudden?* [Physician]
>
> *No, I don't really play football, just in the breaks...[Oliver]*

The physician returned to this issue later during the consultation:

> *Then what about getting in better shape. Besides playing a bit of football in the schoolyard, is there something specific you do?* [Physician]
>
> *Well, I actually think I'm in OK shape.* [Oliver]
>
> *Right OK, but I think we've talked about it a couple of times.* [Physician]

By referring to the patient's own statement, the physician can ask the question:

> *Then what about getting in better shape [...] is there something specific you do?*

Hence, the physician tries to *correct* the deviation, that is, encourage the patient to improve his physical condition. This type of questioning continues and included the nurse who asked whether the patient participates in physical education. Oliver's mother showed compliance and mentioned that they have talked about taking long walks during weekends and signing up for membership at a fitness center. The physician commented "Well, it sounds like you have a lot of good things up and running." Again, compliant behavior was encouraged.

How-R-you as a Power Technology

Our findings suggest that three interdependent practices constitute How-R-you as a power technology: first, it orders patients' words and presents them in a specific way. Second, it structures the patients' speech during conversations around deviations. In doing so, it may have a silencing effect on patients' voices. Third, it allows for the disciplining of patients by referring to their own words and speech, thereby enabling an exercise of power that does not appear oppressive. Thus, these practices are means by which disciplinary power is exercised, which entices, motivates, and commends the compliant behavior, and at the same time constructs noncompliant behavior as *wrong*. Naturally, we do not argue that specific suggestions in the discipline of patients are wrong or damaging. However, we argue that compliant and deviant behaviors are determined based on medical knowledge and expertise, and patients' voices are not included in this assessment. In Foucauldian terms, the exercise of power might produce more healthy subjects observed through a medical gaze focused on lifestyle choices, but in doing so, it might ignore relevant aspects of the patients' lives. This conflicts with the purpose of empowerment in chronic disease self-management, which emphasizes the strengthening of patients' voices and the importance of patients' everyday lives with the disease. As a power technology, How-R-you is an efficient management tool that spares physicians from lengthy conversations, which are often affected by patients' recall bias. Instead, it presents PRO data in a structured manner that allows physicians to quickly identify deviations. In a Foucauldian sense, the deviations become the object of a power that *corrects* behavior. It is therefore questionable whether How-R-you supports patient empowerment during consultations. Our analysis revealed practices in which patients' behaviors were governed toward specific ends, which might result in compliance. However, as a result, their voices may not be heard, and thus, relevant aspects of their lives may be ignored. In the following section, we discuss the implications of these results.

Discussion

Contribution and Implications

First, our results are surprising because they are in conflict with the perceived purpose of mHealth technology as described in the literature, which defines empowerment as giving patients greater control of their health and disease treatment [8,31,32]. Moreover, the findings are particularly interesting because they reveal a dimension to the study and use of mHealth technology, which has so far been largely neglected by the extant literature. Studies emphasize the importance of including patients in technology design [1,33,36] and studying the patient-physician relationship through data collected before, between, and after consultations [19]. Although How-R-you is developed through a user-centered design process, our study shows that the use of technology constitutes practices that promote compliance. This means that although user-centered design is a prerequisite for empowerment, it is not sufficient. The actual use of technologies during consultations also affects empowerment. In this particular case study, there is evidence to suggest that the technology could have a disciplinary rather than empowering effect during consultations, as it specifically structures the consultation around deviations and compliance. Our qualitative research design, with its focus on consultations from a process perspective, allows us to gain insights into how a particular technology orders and structures the consultation and patient-physician relationship and what the implications are for patient empowerment. Thus, we contribute to state-of-the-art knowledge by showing that technology use may have negative effects on empowerment, and we underscore the importance of future research to investigate the actual uses of different technologies.

From the perspective of the physician, our results show how technologies such as How-R-you construct records of patient registrations of PRO data, which are subsequently used to structure conversations with patients and discipline them by referring to their own words and thus foster compliance. This discipline occurs through education, commending compliant behavior, and questioning deviations. The disciplining of patients may also take the form of, for example, repeating patients' own words that indicate a desire to change behavior toward compliance [42]. Disciplining in this sense is about making use of patients' own words and statements to make them recognize a willingness to comply with treatment plans and medical recommendations. Physicians may use technologies during consultations to improve medical compliance, and our results show the manner in which this discipline occurs. Physicians may also use our study to increase their awareness of how technologies structure their observations. In doing so, they may come to appreciate that information other than that which the technology presents as *deviant* may also be relevant to the patients' lives and thus for their ability to self-manage their disease.

By contrast, to assert control, patients need to become aware of practices reducing them to mere deviations that must be *corrected*. Thereby, patients can take advantage of the empowering potential of technologies to steer conversations toward registrations that are important to them. For instance, patients could direct attention to registrations that might have gone unnoticed by the physician and expand on these registrations by dwelling on details that are important to them. Thus, as our findings underscore, this empowering potential does not reside innately in technologies. To nurture this potential, both physicians and patients must become aware of the way in which technologies organize and structure their relationships. Practical means of doing so include physicians spending time examining the registrations with the patients. In doing so, not only should deviations and noncompliance be attended to, but the physician should also initiate dialogue around what may at first glance seem like unremarkable registrations, asking questions such as, "How did you feel during this time, when you were seemingly physically active?" and "Did you do or notice something different?" Thereby, patients are encouraged to speak not only about reasons for deviations or noncompliance but also about other aspects that are relevant to their lives with the disease.

In conclusion, we argue that mHealth technology cannot simply be assumed to be empowering but must be studied in the specific contexts in which it is used. On one hand, it may effectively reduce the recall bias and provide physicians with more data. On the other hand, while studying empowerment from a patient perspective, it cannot be disentangled from its specific uses during consultations. Empowerment depends on patient-physician relationships, where data are shared and where patients' voices and experiences are heard. This has implications for the design of mHealth technologies. Given that empowerment depends on allowing patients' voices and their specific experiences to be heard, designers should prioritize the development of features that support dialogue around specific patient experiences. For instance, instead of turning quantitative PRO data into graphs, the technology may allow patients to register experiences for follow-up and elaboration during consultations. Overall, in addition to emphasizing the collection and sharing of data, mHealth technologies could incorporate design features that aim to invoke dialogue between patients and physicians where the patients' lived experiences with the disease are heard.

Call for Research

First, this study raises the question of whether other types of mHealth technology used in chronic disease management have similar effects. As mentioned, in extant literature, the question of how empowerment happens (or not) during consultations is understudied. Our study underscores the importance of studying how technology impacts patient-physician consultation to understand its potential to empower patients. We neither intend nor are we able to reveal a general pattern in the use of mHealth technologies during consultations. For generalization, additional cases need to be investigated, including different technologies, practices, and patient groups. Second, research is also needed to understand how the identified practices influence the work practices of physicians. Foucault emphasizes that power is not exercised by subject A over subject B. Power is relational and thus affects both subjects. Technologies that privilege patients' autonomy challenge physicians' reliance on medical knowledge. The question is, therefore, how physicians redefine themselves when they are expected to manage without managing? Third, our results raise questions regarding the relationship between empowerment and compliance, which calls for future investigations. Is the former always preferable to the latter, or is it possible to achieve both at the same time without compliance, preventing patients from assuming control and vice versa? Furthermore, it is possible that practices similar to those identified in this study lead to patients feeling empowered despite being disciplined. This raises questions about *real* versus perceived empowerment. To what extent does perceived empowerment improve clinical outcomes, and if it involves disciplining the patients, to what extent does it support patients in dealing with their diseases?

Limitations

A number of limitations of our study deserve mention. First, in our analysis, we did not differentiate between patient ages. For our analytical purposes, we focused on how How-R-you structures consultations. However, we recognize that physicians may approach consultations with patients from different age groups differently, and that different age groups react differently to mHealth technology and medical advice. Second, as the patients in our study were children and adolescents, the way How-R-you structures the consultations may vary from that of adults, and power relations will take different forms. Although we recognize that consultations with children and adults differ, we believe that the questions concerning mHealth and power relations raised by our study apply to both types of consultations. Thus, while our study does not make these differentiations, it points to the importance of studying the use of mHealth in different contexts and patient groups.

Conclusions

In this paper, we investigate the use of mHealth technology during patient-physician consultations with a focus on empowerment. We thereby shed light on an area that remains understudied, namely the use of mHealth during consultations. On the basis of our analytical findings, we argue that the use of the How-R-you app constitutes it as a power technology that promotes compliance with medical recommendations. Our findings contribute to the extant literature by showing how practices, emanating from the use of mHealth technologies, may have unintended consequences. This calls for future research into the use of various types of mHealth technology during consultations to investigate whether they in fact support patients in asserting control of their health and disease treatment.

Acknowledgments

The authors would like to thank the physicians, patients, and parents for their participation. Their contributions made this study possible.

Conflicts of Interest

None declared.

References

1. Dadgar M, Joshi K. The role of information and communication technology in self-management of chronic diseases: an empirical investigation through value sensitive design. J Assoc Inf Syst 2018 Feb;19(2):86-112. [doi: 10.17705/jais1.00485]

2. Lasorsa I, D'Antrassi P, Ajčević M, Stellato K, Di Lenarda A, Marceglia S, et al. Personalized support for chronic conditions. Appl Clin Inform 2017 Dec 19;07(03):633-645. [doi: 10.4338/aci-2016-01-ra-0011]
3. Noncommunicable diseases. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2021-05-25]
4. Noncommunicable diseases and their risk factors. World Health Organization. URL: https://www.who.int/ncds/en/ [accessed 2021-05-25]
5. Honka A, Kaipainen K, Hietala H, Saranummi N. Rethinking health: ICT-enabled services to empower people to manage their health. IEEE Rev Biomed Eng 2011;4:119-139. [doi: 10.1109/rbme.2011.2174217]
6. The NHS long term plan. National Health Service. 2019. URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf [accessed 2021-05-25]
7. Nationale Mål for Sundhedsvæsenet. SUM [Danish Ministry of Health and for the Elderly]. 2018. URL: https://www.regioner.dk/media/9769/nationalemaal_2018.pdf [accessed 2021-05-25]
8. Macdonald GG, Townsend AF, Adam P, Li LC, Kerr S, McDonald M, et al. eHealth technologies, multimorbidity, and the office visit: qualitative interview study on the perspectives of physicians and nurses. J Med Internet Res 2018 Dec 26;20(1):e31 [FREE Full text] [doi: 10.2196/jmir.8983] [Medline: 29374004]
9. Barr P, Scholl I, Bravo P, Faber MJ, Elwyn G, McAllister M. Assessment of patient empowerment - a systematic review of measures. PLoS One 2015;10(5):e0126553 [FREE Full text] [doi: 10.1371/journal.pone.0126553] [Medline: 25970618]
10. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res 2013 Feb 20;15(2):e37. [doi: 10.2196/jmir.2281]
11. Ong S, Jassal SV, Porter E, Logan AG, Miller JA. Using an electronic self-management tool to support patients with chronic kidney disease (CKD): a CKD clinic self-care model. Semin Dial 2013;26(2):195-202. [doi: 10.1111/sdi.12054] [Medline: 23406283]
12. Motamarri S, Akter S, Ray P, Tseng C. Distinguishing "mHealth" from other healthcare services in a developing country: a study from the service quality perspective. Commun Assoc Inf Syst 2014;34:669-692. [doi: 10.17705/1cais.03434]
13. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
14. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. US Food & Drug Administration. 2009. URL: https://tinyurl.com/2cz635uj [accessed 2021-05-25]
15. Rivera SC, Kyte DG, Aiyegbusi OL, Slade AL, McMullan C, Calvert MJ. The impact of patient-reported outcome (PRO) data from clinical trials: a systematic review and critical analysis. Health Qual Life Outcomes 2019 Oct 16;17(1):156 [FREE Full text] [doi: 10.1186/s12955-019-1220-z] [Medline: 31619266]
16. Müller S, Mahler B, Mathiasen M. The value of patient-reported outcome (PRO) data in digital healthcare: using the How-R-You app as a PRO instrument. Hawaii Int Conf Sys Sci (HICSS) 2019:3905-3914. [doi: 10.24251/hicss.2019.473]
17. Steele GC, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
18. Risling T, Martinez J, Young J, Thorp-Froslie N. Defining empowerment and supporting engagement using patient views from the citizen health information portal: qualitative study. JMIR Med Inform 2018 Sep 10;6(3):e43 [FREE Full text] [doi: 10.2196/medinform.8828] [Medline: 30201603]
19. Bradway M, Giordanengo A, Joakimsen R, Hansen AH, Grøttland A, Hartvigsen G, et al. Measuring the effects of sharing mobile health data during diabetes consultations: protocol for a mixed method study. JMIR Res Protoc 2020 Feb 10;9(2):e16657 [FREE Full text] [doi: 10.2196/16657] [Medline: 32039818]
20. Petrič G, Atanasova S, Kamin T. Impact of social processes in online health communities on patient empowerment in relationship with the physician: emergence of functional and dysfunctional empowerment. J Med Internet Res 2017 Mar 13;19(3):e74 [FREE Full text] [doi: 10.2196/jmir.7002] [Medline: 28288953]
21. Mathieu-Fritz A, Guillot C. Diabetes self-monitoring devices and transformations in "patient work". New forms of temporality, reflexivity and self-knowledge relating to the experience of chronic illness. Revue D'anthropologie Des Connaossances 2017;11:k à ao [FREE Full text] [doi: 10.3917/rac.037.k]
22. Liu X, Zhang B, Susarlia A, Padman R. Go to you tube and call me in the morning: use of social media for chronic conditions. MIS Q 2020 Jan 01;44(1):257-283. [doi: 10.25300/MISQ/2020/15107]
23. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
24. Pare G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007 May 01;14(3):269-277. [doi: 10.1197/jamia.m2270]

25. Wootton R, Blignault I, Cignoli J. A national survey of telehealth activity in Australian hospitals. J Telemed Telecare 2003 Dec 02;9 Suppl 2(2_suppl):73-75. [doi: 10.1258/135763303322596345] [Medline: 14728769]
26. Ekeland A, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
27. Frøisland DH, Årsand E. Integrating visual dietary documentation in mobile-phone-based self-management application for adolescents with type 1 diabetes. J Diabetes Sci Technol 2015 May;9(3):541-548 [FREE Full text] [doi: 10.1177/1932296815576956] [Medline: 25901020]
28. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 04;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
29. Waite-Jones JM, Majeed-Ariss R, Smith J, Stones SR, Van RV, Swallow V. Young people's, parents', and professionals' views on required components of mobile apps to support self-management of juvenile arthritis: qualitative study. JMIR Mhealth Uhealth 2018 Jan 19;6(1):e25 [FREE Full text] [doi: 10.2196/mhealth.9179] [Medline: 29351898]
30. 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;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
31. Calvillo J, Román I, Roa LM. How technology is empowering patients? A literature review. Health Expect 2015 Oct;18(5):643-652 [FREE Full text] [doi: 10.1111/hex.12089] [Medline: 23711169]
32. Reychav I, Parush A, McHaney R, Hazan M, Moshonov R. The use of mobile technology in waiting rooms to leverage women's empowerment: a conceptual context. Health Informatics J 2018 Sep 13;24(3):277-292 [FREE Full text] [doi: 10.1177/1460458216671561] [Medline: 27738260]
33. Storni C. Design challenges for ubiquitous and personal computing in chronic disease care and patient empowerment: a case study rethinking diabetes self-monitoring. Pers Ubiquit Comput 2013 Aug 14;18(5):1277-1290 [FREE Full text] [doi: 10.1007/s00779-013-0707-6]
34. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with ehealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
35. Fioravanti A, Fico G, Salvi D, García-Betances RI, Arredondo MT. Automatic messaging for improving patients engagement in diabetes management: an exploratory study. Med Biol Eng Comput 2015 Dec;53(12):1285-1294. [doi: 10.1007/s11517-014-1237-8] [Medline: 25564181]
36. Butterworth K, Allam O, Gray A, Butterworth H. Providing confusion: the need for education not information in chronic care. Health Informatics J 2012 Jun 24;18(2):111-123 [FREE Full text] [doi: 10.1177/1460458212445500] [Medline: 22733679]
37. Kishik S, Müller S, Stampe K. Improving the patient-physician dialogue through Health Information Systems (HIS): misconceptions and miscommunication. Hawaii Int Conf Sys Sci 2020 2020:3460-3469. [doi: 10.24251/HICSS.2020.424]
38. Petrakaki D, Hilberg E, Waring J. Between empowerment and self-discipline: governing patients' conduct through technological self-care. Soc Sci Med 2018 Sep;213:146-153 [FREE Full text] [doi: 10.1016/j.socscimed.2018.07.043] [Medline: 30081356]
39. Keshet Y, Popper-Giveon A. The undisciplined patient in neoliberal society: conscious, informed and intuitive health behaviours. Health Risk Soc 2018 Jan 29;20(3-4):182-199. [doi: 10.1080/13698575.2018.1432757]
40. Foucault M. The subject and power. Crit Inq 1982 Jul;8(4):777-795. [doi: 10.1086/448181]
41. Foucault M. Discipline and Punish: The Birth of the Prison. London: Penguin; 1977.
42. Karlsen MP, Villadsen K. Who should do the talking? The proliferation of dialogue as governmental technology. Cult Organ 2008 Dec;14(4):345-363. [doi: 10.1080/14759550802489680]
43. Kendall G, Michael M. Order and disorder: time, technology and the self. Culture Machine. URL: https://culturemachine.net/interzone/order-and-disorder-kendall-michael/ [accessed 2021-05-25]
44. Coveney J. The government and ethics of health promotion: the importance of Michel Foucault. Health Educ Res 1998;13(3):459-468 [FREE Full text] [doi: 10.1093/her/13.3.459]
45. Rose N. Governing the Soul: The Shaping of the Private Self. London: Routledge; 1990:1-352.
46. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002 Oct;48(2):177-187. [doi: 10.1016/S0738-3991(02)00032-0]
47. Stampe K. Digitally Enabled Chronic Disease Self-Management: Embodiment of Digital Artifacts in JIA Patients' Everyday Life. Aarhus: Aarhus University, Department of Management; 2020.
48. Malinowski B. A Diary in the Strict Sense of the Word. Milton, UK: Routledge; 1967.
49. Whyte W. Street Corner Society: The Social Structure of an Italian Slum. Chicago: The University of Chicago Press; 1993.
50. O'Reilly K. Ethnographic Methods. 2nd ed. London - New York: Routledge; 2012.
51. Geertz C. Thick description: toward an interpretive theory of culture. In: Geertz C, editor. The Interpretation Of Cultures: Selected Essays. New York: Basic Books; 1975:3-30.
52. Spradley JP. Participant Observation. Wadsworth: Thomson Learning; 1980.

53. Tong A, Jones J, Craig JC, Singh-Grewal D. Children's experiences of living with juvenile idiopathic arthritis: a thematic synthesis of qualitative studies. Arthritis Care Res (Hoboken) 2012 Sep;64(9):1392-1404. [doi: 10.1002/acr.21695] [Medline: 22504867]
54. McGrath P, Seifert C, Speechley K. A new analogue scale for assessing children’s pain: an initial validation study. Pain 1996;64:435-443. [doi: 10.1016/0304-3959(95)00171-9]
55. Shields B, Palermo T, Powers J, Grewe SD, Smith GA. Predictors of a child's ability to use a visual analogue scale. Child Care Health Dev 2003 Jul;29(4):281-290 [FREE Full text] [doi: 10.1046/j.1365-2214.2003.00343.x] [Medline: 12823333]
56. Beales JG, Holt PJ, Keen JH, Mellor VP. Children with juvenile chronic arthritis: their beliefs about their illness and therapy. Ann Rheum Dis 1983 Oct 01;42(5):481-486 [FREE Full text] [doi: 10.1136/ard.42.5.481] [Medline: 6625698]
57. De Monte R, Rodger S, Jones F, Broderick S. Living with juvenile idiopathic arthritis: children's experiences of participating in home exercise programmes. Br J Occup Ther 2009 Aug 01;72(8):357-365. [doi: 10.1177/030802260907200806]
58. AAA Ethics Forum. Principles of professional responsibility. 2012. URL: http://ethics.americananthro.org/category/statement/ [accessed 2021-05-25]

Abbreviations

JIA: juvenile idiopathic arthritis
mHealth: mobile health
PRO: patient-reported outcome

Edited by R Kukafka; submitted 06.01.21; peer-reviewed by J Waite-Jones, G Macdonald, L Alpay; comments to author 23.02.21; revised version received 03.04.21; accepted 06.05.21; published 15.06.21.

Please cite as:
Stampe K, Kishik S, Müller SD
Mobile Health in Chronic Disease Management and Patient Empowerment: Exploratory Qualitative Investigation Into Patient-Physician Consultations
J Med Internet Res 2021;23(6):e26991
URL: https://www.jmir.org/2021/6/e26991
doi:10.2196/26991
PMID:34128817

Original Paper

Patient Experience in Virtual Visits Hinges on Technology and the Patient-Clinician Relationship: A Large Survey Study With Open-ended Questions

Susannah Rose[1*], PhD; Heather McKee Hurwitz[2*], PhD; Mary Beth Mercer[1*], MPH; Sabahat Hizlan[1*], BA, BS; Kari Gali[3*], DNP; Pei-Chun Yu[4*], MS; Caroline Franke[1*], MD; Kathryn Martinez[5*], PhD; Matthew Stanton[3*], MBA; Matthew Faiman[3], MD; Peter Rasmussen[3*], MD; Adrienne Boissy[1*], MA, MD

[1]Office of Patient Experience, Clinical Transformation, Cleveland Clinic, Cleveland, OH, United States
[2]Taussig Cancer Center, Cleveland Clinic, Cleveland, OH, United States
[3]Digital Health, Cleveland Clinic, Cleveland, OH, United States
[4]Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, United States
[5]Center for Value-Based Care Research, Cleveland Clinic, Cleveland, OH, United States
[*]these authors contributed equally

Corresponding Author:
Susannah Rose, PhD
Office of Patient Experience, Clinical Transformation
Cleveland Clinic
9500 Euclid Avenue, JJ60
Cleveland, OH, 44195
United States
Phone: 1 216 444 0849
Email: roses2@ccf.org

Abstract

Background: Patient satisfaction with in-person medical visits includes patient-clinician engagement. However, communication, empathy, and other relationship-centered care measures in virtual visits have not been adequately investigated.

Objective: This study aims to comprehensively consider patient experience, including relationship-centered care measures, to assess patient satisfaction during virtual visits.

Methods: We conducted a large survey study with open-ended questions to comprehensively assess patients' experiences with virtual visits in a diverse patient population. Adults with a virtual visit between June 21, 2017, and July 12, 2017, were invited to complete a survey of 21 Likert-scale items and textboxes for comments following their visit. Factor analysis of the survey items revealed three factors: experience with technology, patient-clinician engagement, and overall satisfaction. Multivariable logistic regression was used to test the associations among the three factors and patient demographics, clinician type, and prior relationship with the clinician. Using qualitative framework analysis, we identified recurrent themes in survey comments, quantitatively coded comments, and computed descriptive statistics of the coded comments.

Results: A total of 65.7% (426/648) of the patients completed the survey; 64.1% (273/426) of the respondents were women, and the average age was 46 (range 18-86) years. The sample was geographically diverse: 70.2% (299/426) from Ohio, 6.8% (29/426) from Florida, 4.2% (18/426) from Pennsylvania, and 18.7% (80/426) from other states. With regard to insurance coverage, 57.5% (245/426) were undetermined, 23.7% (101/426) had the hospital's employee health insurance, and 18.7% (80/426) had other private insurance. Types of virtual visits and clinicians varied. Overall, 58.4% (249/426) of patients had an on-demand visit, whereas 41.5% (177/426) had a scheduled visit. A total of 41.8% (178/426) of patients had a virtual visit with a family physician, 20.9% (89/426) with an advanced practice provider, and the rest had a visit with a specialist. Most patients (393/423, 92.9%) agreed that their virtual visit clinician was interested in them as a person, and their virtual visit made it easy to get the care they needed (383/421, 90.9%). A total of 81.9% (344/420) of respondents agreed or strongly agreed that their virtual visit was *as good as* an in-person visit by a clinician. Having a prior relationship with their virtual visit clinician was associated with less comfort and ease with virtual technology among patients (odds ratio 0.58, 95% CI 0.35-0.98). In terms of technology, patients found the interface easy to use (392/423, 92.7%) and felt comfortable using it (401/423, 94.8%). Technical difficulties were associated with lower odds of overall satisfaction (odds ratio 0.46, 95% CI 0.28-0.76).

Conclusions: Patient-clinician engagement in virtual visits was comparable with in-person visits. This study supports the value and acceptance of virtual visits. Evaluations of virtual visits should include assessments of technology and patient-clinician engagement, as both are likely to influence patient satisfaction.

(J Med Internet Res 2021;23(6):e18488) doi:10.2196/18488

KEYWORDS

telehealth; virtual visit; patient experience; mobile phone

Introduction

Virtual visits have emerged at a time in health care when clinicians' decisions have become only one part of patients' experiences. Patients' and their families' needs, values, and inputs influence care decisions [1]. Patients are researching treatments on the internet and engaging as advocates and even as partners with physicians both on the internet and in person [1-5]. A crucial aspect of patient experience is the patient-clinician engagement or using a *relationship-centered care* approach. This approach prioritizes how the communication and emotions shared between patients and clinicians gradually affect patients' experiences as a whole person [5,6]. Furthermore, whether physicians empathize with their patients strongly shapes the patient-physician relationship and patient satisfaction [7]. Virtual visits must continue the best aspects of in-person care to be an acceptable substitute for in-person care. However, no studies have evaluated whether virtual visits are addressing patients' experiences comprehensively rather than simply whether or not they are satisfied.

Studies on virtual visits have examined patient satisfaction. For example, patients have reported positive experiences about communication, engagement, and building rapport with clinicians during their virtual visits [8-12]. Furthermore, patients have expressed that virtual visits are convenient in decreasing travel time, wait time, and stress and increasing comfort and convenience [13-15]. Virtual visits increased substantially during the initial phase of the COVID-19 pandemic, and physicians reported that patients respond positively to virtual visits [16], that telemedicine addresses both urgent and nonurgent care [17], and that telehealth has become a valuable way to evaluate patients before they enter a medical facility, thereby limiting potential clinician or patient exposure to COVID-19 [18]. Despite the increasing use of telehealth during the COVID-19 pandemic, these studies have not evaluated whether virtual visits actually serve the patient-clinician relationship [16-18]. In addition, studies that have examined patient satisfaction with virtual visits are mostly specialty- or disease-specific [19-21], have small sample sizes or low survey response rates [9,10,21,22], or include only one item to measure patient satisfaction [13].

To gain a more comprehensive understanding of the patient experience during virtual visits, especially patient-clinician engagement in a range of types of virtual visits, we conducted a survey study with patients who received virtual care from a variety of clinicians representing different practices.

Methods

We conducted a large survey study that included open-ended questions to comprehensively assess patients' experiences with virtual visits in the Cleveland Clinic Health System. This cross-sectional study was conducted with adults who had a synchronous virtual visit between June 21, 2017, and July 12, 2017. These visits took place on a synchronous video-visit platform developed by the American Well [23]. The study examined scheduled and on-demand virtual visits that used audio and video (telephonic only were not available). With this app, patients can locate and connect with clinicians in an *on-demand* fashion, and clinicians can also schedule visits in advance so that patients can simply click on a link at the designated time to connect directly with their clinician. Visits were termed an *Express Care Online* visit because patients could access a clinician quickly, though still comprehensively receive help. The user interface facilitates virtual visits over smartphones, tablets, and computers [23].

Eligible patients were identified by daily census reports from the previous day's virtual visits. When patients had more than one virtual visit during study accrual, only the first visit was included. In addition, visits used for clinician training, digital health testing, and those with a duration of 0 minutes were excluded from the study. All 648 eligible patients with visits during that period were sent an email invitation to complete a brief web-based survey regarding their experiences with their recent virtual visit. All nonresponders were sent reminder email invitations and were then called by a member of the research team. Patients were given a US $5 Amazon e-gift card upon completion of the survey. As part of their participation, they were informed about the gift card at the time of recruitment. Given the small amount of compensation, participation bias is likely to be minimal.

Survey items were designed to consider patients as whole persons, evaluate the patient-clinician relationship, and broadly assess patients' experiences with virtual visits; specifically, we sought to assess patient's overall satisfaction, overall ease, technical ease, convenience, relationship with the clinician, and perception of clinician empathy. The survey also aimed to evaluate how patients compared virtual visit quality and value with in-person visits. Survey questions were developed after an extensive review of the literature [14,15,24] and several iterations of expert stakeholder input from members of the digital health team, the office of patient experience, and patient volunteers. The survey consisted of 21 items evaluated on a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree) and text boxes for comments on what patients liked best about their virtual visit and what could be improved. The survey items are presented in Multimedia Appendix 1.

Survey data were collected and managed using the REDCap (Research Electronic Data Capture) electronic data capture system [25]. Survey data were merged with data from the hospital system that included virtual visit characteristics, such as visit type (scheduled or on-demand), device type (computer or mobile device), clinician type (physician or advanced practice provider), clinician specialty, insurance type, and patient demographic information.

First, we computed descriptive statistics. Factor analysis of the 21 survey items revealed three factors: overall satisfaction with the virtual visit (standardized Cronbach α=.93), comfort and ease using virtual visit technology (Cronbach α=.89), and patient-clinician engagement (Cronbach α=.92). Items associated with each factor were summed and dichotomized at the mean for analysis. We used multivariable logistic regression with backward variable selection to assess the associations among patient demographics; virtual visit clinician type and prior patient relationship with their virtual visit clinician; and overall satisfaction with their virtual visit, patient-reported experience using the virtual visit technology, and patient-clinician engagement. Quantitative data were analyzed using SAS 9.4 [26].

For the open-ended questions, a framework analysis of the content was used to identify themes in patients' narrative comments. Themes were used to create a coding checklist consisting of dichotomous (yes or no) variables. Themes that emerged from comments in response to what patients appreciated about their virtual visit included not having to travel outside home to receive care, enhanced access to care, and convenience. Themes related to recommendations for improvements included enhancements to the user interface and instructions to better prepare for virtual visits. A coding checklist was used to code the text data quantitatively and descriptive statistics were computed.

This study was approved by our medical center's institutional review board.

Results

Overview

Out of the 426 patients who responded to the survey (426/648, 65.7%), nearly two-thirds self-identified as female (273/426, 64.1%). The average age was 46 (range 18-86) years; 70.2% (299/426) of the patients were from Ohio, 6.8% (29/426) were from Florida, 4.2% (18/426) were from Pennsylvania, and 18.7% (80/426) resided in other states. Insurance coverage was undetermined for more than half of the patients (245/426, 57.5%), 23.7% (101/426) had the hospital's employee health insurance, and 18.7% (80/426) had other private insurance. The demographic characteristics of the respondents were similar to those of the nonrespondents.

A total of 58.4% (249/426) patients had an on-demand virtual visit for an acute concern, and 41.5% (177/426) had a scheduled virtual visit; 28.6% (121/423) of patients knew their virtual visit clinician from a previous in-person appointment with that same clinician, and those patients were older (*P*=.003) and more often male (*P*=.03) compared with patients who did not know their virtual visit clinician. Moreover, most patients who knew their virtual visit clinician had a scheduled rather than on-demand virtual visit (119/121, 98.3%). About 41.8% (178/426) of clinicians were family physicians, whereas the rest represented other physician specialties, and 20.9% (89/426) were advanced practice providers. Most patients used mobile devices such as phones or tablets for their visits (348/424, 82.1%), whereas the rest were connected using a computer. Patient and virtual visit characteristics are presented in Table 1.

Table 1. Patient and virtual visit characteristics (N=426).

Characteristics	Value
Age (years), mean (SD)	46 (15.5)
Female, n (%)	273 (64.1)
Ohio resident, n (%)	299 (70.2)
Hospital employee, n (%)	101 (23.7)
Insurance used, n (%)	
Employee Health Plan	101 (23.7)
Medical Mutual of Ohio	80 (18.8)
Undetermined	245 (57.5)
Physician virtual clinician, n (%)	336 (78.9)
Prior relationship with clinician (n=423), n (%)	121 (28.6)
Visit type, n (%)	
On-demand	249 (58.4)

Overall Satisfaction

The mean overall satisfaction score was 4.4 out of 5 with an SD of 0.78 (see Table 2 for detailed Likert scale survey results). Overall, 81.9% (344/420) of respondents agreed (107/420, 25.5%) or strongly agreed (237/420, 56.4%) that their virtual visit was *as good as* an in-person visit with a clinician. More than half of the respondents agreed (73/417, 17.5%) or strongly agreed (149/417, 35.7%) that their virtual visit was *better than* an in-person visit with a clinician. In multivariable logistic regression, employee patient status (vs nonemployee) was associated with higher odds of overall satisfaction with their virtual visit (odds ratio [OR] 1.9, 95% CI 1.14-3.47). The results from the multivariable logistic regression are presented in Table 3.

Table 2. Patient responses to Likert scale survey items (N=426).

Item	Total, n	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
Overall satisfaction						
My virtual visit made it easy to get the care I needed.	421	6 (1.4)	14 (3.3)	18 (4.3)	94 (22.3)	289 (68.6)
For my health concern, my virtual visit was *as good as* an in-person visit with a health care provider.	420	9 (2.1)	27 (6.4)	40 (9.5)	107 (25.5)	237 (56.4)
For my health concern, my virtual visit was *better than* an in-person visit with a health care provider.	417	19 (4.6)	49 (11.7)	127 (30.5)	73 (17.5)	149 (35.7)
My virtual visit saved me time.	422	9 (2.1)	6 (1.4)	19 (4.5)	62 (14.7)	326 (77.2)
My virtual visit was worth the money I spent on the visit.	420	21 (5)	8 (1.9)	46 (10.9)	81 (19.3)	264 (62.9)
I would use a virtual visit again.	422	8 (1.9)	7 (1.7)	15 (3.5)	90 (21.3)	302 (71.6)
I would recommend a virtual visit to others.	414	9 (2.2)	8 (1.9)	17 (4.1)	82 (19.8)	298 (72)
Comfort and ease using virtual visit technology						
My virtual visit platform was easy to use.	422	3 (0.7)	13 (3.1)	12 (2.8)	103 (24.4)	291 (68.9)
I was comfortable using my virtual visit platform.	423	0 (0)	2 (0.5)	20 (4.7)	105 (25.8)	296 (70)
The wait time to see my online health care provider was reasonable.	422	3 (0.7)	4 (0.9)	15 (3.6)	107 (25.4)	293 (69.4)
It was easy to see my health care provider during my online visit.	424	7 (1.6)	5 (1.2)	9 (2.1)	91 (21.5)	312 (73.6)
It was easy to hear my health care provider during my online visit.	423	9 (2.1)	9 (2.1)	11 (2.6)	92 (21.8)	302 (71.4)
It was easy to talk with my health care provider during my online visit.	423	7 (1.6)	6 (1.4)	10 (2.5)	87 (20.6)	313 (74)
The technology was easy to use.	423	5 (1.2)	8 (1.9)	18 (4.3)	105 (24.8)	287 (67.8)
Patient-clinician engagement						
My online health care provider was interested in me as a person.	423	2 (0.5)	2 (0.5)	26 (6.1)	98 (23.2)	295 (69.7)
My online health care provider fully understood my health concern.	417	0 (0)	4 (0.9)	14 (3.4)	86 (20.6)	313 (75.1)
My online health care provider and I made a plan of action to resolve my health concern.	421	2 (0.5)	6 (1.4)	15 (3.6)	95 (22.6)	303 (72)
I believe that the plan of action my health care provider recommended will resolve my health concern.	419	6 (1.4)	8 (1.9)	54 (12.9)	89 (21.2)	262 (62.5)
I understand what I need to do next to resolve my health concern.	422	1 (0.2)	3 (0.7)	18 (4.3)	106 (25.1)	294 (69.7)
I had enough time with my health care provider during my online visit.	418	1 (0.2)	4 (0.9)	14 (3.3)	102 (24.4)	297 (71)
My privacy was respected during my online visit.	415	0 (0)	0 (0)	10 (2.4)	102 (24.6)	303 (73)

Table 3. Predictors of overall satisfaction, patient-clinician engagement, and comfort and ease with virtual technology.

Factor	Odds ratio (95% CI)
Patient-clinician engagement	
Virtual clinician was nurse practitioner or physician assistant (vs family physician)	2.28 (1.25-4.16)
Employee patient status (vs nonemployee)	1.73 (1.01-2.95)
Overall satisfaction	
Employee patient status (vs nonemployee)	1.9 (1.14-3.17)
Technical difficulties (vs no technical difficulties)	0.46 (0.28-0.76)
Comfort and ease using virtual technology	
Prior relationship with virtual clinician (vs no prior relationship)	0.58 (0.35-0.98)

Most patients (383/421, 90.9%) reported that their virtual visit made it easy to receive the care they needed. A respondent noted, “It was quick and easy. Instead of finding babysitters for my 4-month-old twins, it was convenient to do right from my home.” Open-ended responses (Table 4) support this finding: patients valued receiving care from their home without traveling to their clinician’s office (145/363, 39.9%) and reported that virtual visits provided convenient access to health care (121/363, 33.3%) and saved them time (98/363, 26.9%). Most patients reported that they would use virtual care again (392/422, 92.9%) and would recommend it to other people (380/414, 91.8%).

Table 4. Patient feedback on their virtual visit in their own words.

Domain	Patients, n (%)	Illustrative quotes
What did you like best about your virtual visit? (n=363)		
No travel or stay at home	145 (39.9)	“It was so much better to be able to not have to leave the house when I felt so awful and in so much pain.” [P46]
Access or convenience	121 (33.3)	“I used it on a Holiday when my doctor’s office was closed, and it saved me the trip to Urgent Care while in pain.” [P487]
Saved time	98 (26.9)	“I saved time, money and my health by being able to have this visit online.” [P109]
Easy	94 (25.9)	“Having two kids, it was so easy to face time a provider to get the help I needed.” [P99]
Quick	60 (16.5)	“The process was quick.” [P418]
Clinician	50 (13.8)	“My online visit was with Dr. [name] and it was an honest pleasure interacting with her. She seemed legitimately interested and concerned regarding my health matter and at no time did I ever feel awkward or rushed.” [P530]
Cost	36 (9.9)	“It saved me a lot of time and money. My care and concerns were addressed as if I was visiting the doctor in person. I am very pleased.” [P444]
Continuity of care	20 (5.5)	“I loved the fact that I was able to see my provider, and not just anyone. I have seen this doctor for several years, and getting an in-person appointment with her is very hard to do. Therefore, getting an online appointment, I was able to see her faster.” [P82]
What can we improve? (n=100)		
User interface	30 (30)	“I would say the technology was faulty. I’m not an expert, but I followed the directions and it would not connect us.” [P459]
Virtual visit information	14 (14)	“Perhaps a pre-appointment preparation list might help set appropriate expectations for patients. For example, ‘for the following problems, you may need to get additional tests/go to the ER, etc.’ This is because some people may not know these things.” [P96]
Insurance coverage	10 (10)	“I would work to partner with more insurance providers to remove those barriers associated with utilizing health care coverage for this service, if I were the management team.” [P307]
Access to prescriptions	9 (9)	“I was not able to get my prescription for strep through the express care online.” [P70]
Link to the patient portal	5 (5)	“Cross-functionality with MyChart would be great.” [P223]
More clinicians	4 (4)	“I want more specialized care in the app. Please have therapists, psychologists and other specialized doctors.” [P51]

Comfort and Ease Using Virtual Visit Technology

The mean overall technology experience score was 4.6 out of 5 with an SD of 0.57, and 93.4% (394/422) of patients reported that the virtual visit interface was easy to use. Participants agreed and strongly agreed that they felt comfortable using it (401/423, 94.8%), and most agreed or strongly agreed that the wait times for using it were reasonable (400/422, 94.8%). Respondents could see (403/424, 95.1%), hear (394/423, 93.1%), and talk with (400/423, 94.6%) their clinician easily during their virtual visit. However, 19.9% (84/423) of patients reported that they had technical difficulties during their virtual visit (Table 5), and technical difficulties were associated with lower odds of overall satisfaction among patients (OR 0.46, 95% CI 0.28-0.76) in multivariable logistic regression. In addition, in multivariable logistic regression, having a prior relationship with their virtual visit clinician was associated with less comfort and ease with virtual technology among patients (OR 0.58, 95% CI 0.35-0.98). In open-ended comments, 23.5% (100/426) of patients commented on what could be improved, and 30% (30/100) of these comments recommended improvement of the user interface (see Table 4 for additional patient suggestions for improvement).

Table 5. Patient responses to dichotomous survey items (N=426).

Item	Yes; patient, n (%)	No; patient, n (%)
Have you had a previous in-person visit with the health care provider you saw using Express Care Online? (n=423)	121 (28.6)	302 (71.4)
Did you have any technical difficulties during your Express Care Online visit? (n=423)	84 (19.9)	339 (80.1)
Where would you have gone for medical care if you had not used Express Care Online? (n=426)		
Doctor’s Office	229 (53.8)	197 (46.2)
Urgent Care	112 (26.3)	314 (73.7)
Emergency Room	21 (4.9)	405 (95.1)
Retail Clinic	30 (7)	396 (92.9)
I would not have gone for medical care	50 (11.7)	376 (88.3)

Patient-Clinician Engagement

Most patients (393/423, 92.9%) agreed that their virtual visit clinician was interested in them as a person. A respondent commented, “The physician was very kind and really listened to my issue and what I had already tried to resolve it” (for additional patient responses, see Table 4). In addition, the patients felt that they were able to work together with their clinicians reciprocally—94.5% (398/421) of respondents reported that together with their virtual visit clinician, they made a plan of action to resolve their health concern. In multivariable logistic regression, a visit with an advanced practice provider (nurse practitioner or physician assistant) was associated with higher odds of patient-clinician engagement compared with visits with a family medicine physician (OR 2.28, 95% CI 1.25-4.16). In addition, employees of our medical center had higher odds of patient-clinician engagement than nonemployees (OR 1.73, 95% CI 1.01-2.95).

The mean patient-clinician engagement score was 4.6 out of 5, with an SD of 0.53. In open-ended comments, patients described the high quality of their virtual visit clinician (50/363, 13.8%), and 95.7% (399/417) of patients agreed that their virtual visit clinician fully understood their health concern. However, some results suggest a need to improve the patient-clinician relationship. Open-ended comments reflected a need for better orientation to virtual visits: 14% (14/100) of respondents recommended that patients be given more information before their virtual visits to know what to expect and how to prepare for their appointment.

Discussion

Principal Findings

The large and diverse sample size and high response rate (426/648, 65.7%) suggest robust findings about patient experience of virtual visits as compared with prior studies [27].

This large mixed methods study in a major health system found that patients reported high satisfaction with virtual visits, that the technology was easy to use, and that virtual visits were comparable or better than an in-person visit. However, technical difficulties were associated with lower odds of overall satisfaction among patients (OR 0.46, 95% CI 0.28-0.76) in the multivariable logistic regression. Our findings of high patient satisfaction with virtual visits align with those of other published studies [15,20,28,29]. For example, 95% of patients who participated in a MinuteClinic telehealth visit were very satisfied with the quality of the health care they received and rated telehealth as better than or just as good as a traditional in-person visit [15]. However, our findings are unique and go beyond these other studies because we measured satisfaction in terms of several indicators of patient-clinician engagement. We found that most patients in our study reported high engagement with their virtual visit clinician. There are concerns about the effect of telemedicine on trust-based relationships between patients and clinicians [30] and the ability to express empathy in digital settings [31,32]. Our study suggests that it is possible to measure the patient-clinician engagement and begin to evaluate empathy and collaborative relationships with patients during a virtual visit.

However, both clinician and patient identity influenced the likelihood of strong patient-clinician engagement during virtual visits. Having a pre-existing relationship with the clinician was associated with lower satisfaction. In terms of gender and age, older men who had previous relationships with their virtual visit clinicians were some of the patients that were least satisfied with their experiences. Patients who had a scheduled virtual visit were older than those who had an on-demand visit and were perhaps less comfortable with virtual technology than younger patients. In multivariable logistic regression, having a prior relationship with their virtual visit clinician was associated with less comfort and ease with virtual technology among patients (OR 0.58, 95% CI 0.35-0.98). However, patient age was not a significant predictor of overall satisfaction, patient-clinician engagement, or comfort and ease with virtual technology. Patients who received care from their clinician via a traditional in-person appointment may have experienced discomfort with, or simply did not like, this unfamiliar mode of care with a clinician they know. Gender, age, and the necessity of multiple or repeated virtual visits instead of in-person visits may contribute to less satisfaction and should be explored further in future studies.

In terms of employment and insurance status, employee status was associated with overall satisfaction with virtual visits. In multivariable logistic regression, employee patient status (vs nonemployee) was associated with higher odds of overall satisfaction with their virtual visit (OR 1.90, 95% CI 1.14-3.47). In addition, employees of our medical center had higher odds of patient-clinician engagement than nonemployees (OR 1.73, 95% CI 1.01-2.95). This may be because of benefit coverage for the virtual visit; medical center employees received the care they needed with little to no out-of-pocket costs. Employee patients were not charged for their virtual visit, whereas nonemployee patients were charged up to US $49, depending on insurance coverage. Employee patients may also be more familiar with clinicians, telehealth in general, and virtual visits specifically; thus, their expectations may have been better aligned with the virtual health care they received.

Our study has limitations that need to be considered. Although the sample was diverse in age, gender, and visit and clinician type, a large percentage of patients were from Ohio, insurance was unknown for more than half, and a quarter were employees, all of which may limit the generalizability of our findings. In addition, despite our relatively high response rate, we may have a response bias, but this is unlikely given the similarity of the responders to the nonresponders. Furthermore, we should be cautious not to generalize our findings to all virtual visits, given that this was a nonprobability sample in one health system.

In addition, we do not know the reason for patient visits. We collectively analyzed on-demand and scheduled visits together, although on-demand visits are typically low acuity infections (upper respiratory and urinary tract), whereas scheduled visits are usually more complex conditions where patients have more intense needs [13]. As nearly all patients who had a prior relationship with their virtual visit clinician had a scheduled virtual visit rather than an on-demand one, their expectations for the care they would receive were likely different from those of patients who did not know their clinicians. Furthermore, scheduled and on-demand visits may be fundamentally different. This study suggests that virtual visits may be most satisfactory when used for acute problems or when health care access is otherwise limited rather than nonacute or more complex issues when patients may prefer speaking with their provider in person at a medical facility. Although this study provides an analysis of many kinds of visits, future studies should disaggregate the types of visits and analyze them separately.

Future studies should consider both the quality of these visits and medical outcomes because both are likely to influence patient satisfaction.

In addition, future studies should further explore the range of clinician experience in virtual visits. In multivariable logistic regression, a visit with an advanced practice provider (nurse practitioner or physician assistant) was associated with higher odds of patient-clinician engagement than visits with a family medicine physician (OR 2.28, 95% CI 1.25-4.16). Our finding that patients were more likely to report high engagement with advanced practice providers rather than with family physicians aligns with findings on in-person visits [32] and highlights opportunities to leverage advanced practice providers in telehealth. Future studies should explore the basis of these differences.

Future research and efforts should also focus on the user interface, facilitating patient expectations of the technology, and associations with quality.

Conclusions

The impact of the virtual interface on patient-clinician relationships is largely unknown, but our findings are encouraging. Our study found that virtual visits facilitate health care access and relationship-building, contributing to satisfying relationship-centered care, a crucial aspect of contemporary patient experiences. Even during a single virtual visit, we found that patients and clinicians could meaningfully engage in relationship-building practices. Strategies to prepare established patients for virtual visits with their clinicians may ease the transition from in-person care to virtual care, resulting in better experiences for both. Patients should be aware of the capabilities and limitations of patient-clinician engagement in virtual visits [33]. Strategies to prepare clinicians for virtual visits would also support a seamless transition to delivering health care virtually (a tip sheet that outlines 10 best practices for communicating effectively with patients during a virtual visit has been provided in Multimedia Appendix 2) [34].

Conflicts of Interest

PR has a conflict of interest to disclose. He serves as the chief clinical officer of Cleveland Clinic American Well Joint Venture—a digital health company—and receives personal fees for his service on the Strategic Advisory Board of American Well. The other authors have no conflicts of interest.

Multimedia Appendix 1
Express Care Online postvisit patient survey.
[DOCX File , 14 KB - jmir_v23i6e18488_app1.docx]

Multimedia Appendix 2
Top 10 tips for virtual visits: clinician communication.
[DOCX File , 424 KB - jmir_v23i6e18488_app2.docx]

References

1. Lines LM, Lepore M, Wiener JM. Patient-centered, person-centered, and person-directed care: they are not the same. Med Care 2015 Jul;53(7):561-563. [doi: 10.1097/MLR.0000000000000387] [Medline: 26067878]
2. Gerber BS, Eiser AR. The patient physician relationship in the internet age: future prospects and the research agenda. J Med Internet Res 2001;3(2):e15 [FREE Full text] [doi: 10.2196/jmir.3.2.e15] [Medline: 11720957]
3. Lagu T, Norton CM, Russo LM, Priya A, Goff SL, Lindenauer PK. Reporting of patient experience data on health systems' websites and commercial physician-rating websites: mixed-methods analysis. J Med Internet Res 2019 Mar 27;21(3):e12007 [FREE Full text] [doi: 10.2196/12007] [Medline: 30916654]
4. Rosen P, Spalding SJ, Hannon MJ, Boudreau RM, Kwoh CK. Parent satisfaction with the electronic medical record in an academic pediatric rheumatology practice. J Med Internet Res 2011 May 27;13(2):e40 [FREE Full text] [doi: 10.2196/jmir.1525] [Medline: 21622292]
5. Wolf JA, Niederhauser V, Marshburn D, LaVela SL. Reexamining “Defining Patient Experience”: the human experience in healthcare. Patient Exp J 2021 Apr 28;8(1):16-29. [doi: 10.35680/2372-0247.1594]
6. Beach MC, Inui T, Relationship-Centered CR. Relationship-centered care. A constructive reframing. J Gen Intern Med 2006 Jan;21 Suppl 1:3-8 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00302.x] [Medline: 16405707]
7. Eklund JH, Meranius MS. Toward a consensus on the nature of empathy: a review of reviews. Patient Educ Couns 2021 Feb;104(2):300-307 [FREE Full text] [doi: 10.1016/j.pec.2020.08.022] [Medline: 32888755]
8. Mammen JR, Elson MJ, Java JJ, Beck CA, Beran DB, Biglan KM, et al. Patient and physician perceptions of virtual visits for Parkinson's disease: a qualitative study. Telemed J E Health 2018 Apr;24(4):255-267. [doi: 10.1089/tmj.2017.0119] [Medline: 28787250]
9. Agha Z, Schapira RM, Laud PW, McNutt G, Roter DL. Patient satisfaction with physician-patient communication during telemedicine. Telemed J E Health 2009 Nov;15(9):830-839. [doi: 10.1089/tmj.2009.0030] [Medline: 19919189]
10. Salisbury C, Quigley A, Hex N, Aznar C. Private video consultation services and the future of primary care. J Med Internet Res 2020 Oct 01;22(10):e19415 [FREE Full text] [doi: 10.2196/19415] [Medline: 32812887]
11. Bakken S, Grullon-Figueroa L, Izquierdo R, Lee N, Morin P, Palmas W, IDEATel Consortium. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. J Am Med Inform Assoc 2006;13(6):660-667 [FREE Full text] [doi: 10.1197/jamia.M2146] [Medline: 16929036]
12. Elliott T, Tong I, Sheridan A, Lown BA. Beyond convenience: patients' perceptions of physician interactional skills and compassion via telemedicine. Mayo Clin Proc Innov Qual Outcomes 2020 Jun;4(3):305-314 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.04.009] [Medline: 32542222]
13. Martinez KA, Rood M, Jhangiani N, Kou L, Rose S, Boissy A, et al. Patterns of use and correlates of patient satisfaction with a large nationwide direct to consumer telemedicine service. J Gen Intern Med 2018 Oct;33(10):1768-1773 [FREE Full text] [doi: 10.1007/s11606-018-4621-5] [Medline: 30112737]
14. Becevic M, Boren S, Mutrux R, Shah Z, Banerjee S. User satisfaction with telehealth: study of patients, providers, and coordinators. Health Care Manag (Frederick) 2015;34(4):337-349. [doi: 10.1097/HCM.0000000000000081] [Medline: 26506296]
15. Polinski JM, Barker T, Gagliano N, Sussman A, Brennan TA, Shrank WH. Patients' satisfaction with and preference for telehealth visits. J Gen Intern Med 2016 Mar;31(3):269-275 [FREE Full text] [doi: 10.1007/s11606-015-3489-x] [Medline: 26269131]
16. Sasangohar F, Bradshaw MR, Carlson MM, Flack JN, Fowler JC, Freeland D, et al. Adapting an outpatient psychiatric clinic to telehealth during the COVID-19 pandemic: a practice perspective. J Med Internet Res 2020 Oct 01;22(10):e22523 [FREE Full text] [doi: 10.2196/22523] [Medline: 32936768]
17. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Apr 23:1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
18. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Mar 11:1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
19. Kraai IH, Luttik ML, de Jong RM, Jaarsma T, Hillege HL. Heart failure patients monitored with telemedicine: patient satisfaction, a review of the literature. J Card Fail 2011 Aug;17(8):684-690. [doi: 10.1016/j.cardfail.2011.03.009] [Medline: 21807331]

20. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. Br Med J 2000 Jun 3;320(7248):1517-1520 [FREE Full text] [Medline: 10834899]
21. Qiang JK, Marras C. Telemedicine in Parkinson's disease: a patient perspective at a tertiary care centre. Parkinsonism Relat Disord 2015 May;21(5):525-528. [doi: 10.1016/j.parkreldis.2015.02.018] [Medline: 25791380]
22. Williams TL, May CR, Esmail A. Limitations of patient satisfaction studies in telehealthcare: a systematic review of the literature. Telemed J E Health 2001;7(4):293-316. [doi: 10.1089/15305620152814700] [Medline: 11886667]
23. Amwell. American Well Corporation, Boston, MA. URL: https://business.amwell.com/ [accessed 2021-05-22]
24. Nguyen M, Waller M, Pandya A, Portnoy J. A review of patient and provider satisfaction with telemedicine. Curr Allergy Asthma Rep 2020 Sep 22;20(11):72 [FREE Full text] [doi: 10.1007/s11882-020-00969-7] [Medline: 32959158]
25. 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]
26. SAS® 9.4 statements reference. SAS Institute Inc. Cary, NC: SAS Institute Inc; 2013. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.353.6870&rep=rep1&type=pdf [accessed 2021-05-22]
27. Orlando JF, Beard M, Kumar S. Systematic review of patient and caregivers' satisfaction with telehealth videoconferencing as a mode of service delivery in managing patients' health. PLoS One 2019;14(8):e0221848 [FREE Full text] [doi: 10.1371/journal.pone.0221848] [Medline: 31469865]
28. Barnett ML, Ray KN, Souza J, Mehrotra A. Trends in telemedicine use in a large commercially insured population, 2005-2017. J Am Med Assoc 2018 Nov 27;320(20):2147-2149 [FREE Full text] [doi: 10.1001/jama.2018.12354] [Medline: 30480716]
29. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
30. Bean S. Tele-trust: what is telemedicine's impact on the physician-patient relationship? J Clinic Res Bioeth 2015;06(04):e112. [doi: 10.4172/2155-9627.1000e112]
31. Terry C, Cain J. The emerging issue of digital empathy. Am J Pharm Educ 2016 May 25;80(4):58 [FREE Full text] [doi: 10.5688/ajpe80458] [Medline: 27293225]
32. Howick J, Steinkopf L, Ulyte A, Roberts N, Meissner K. How empathic is your healthcare practitioner? A systematic review and meta-analysis of patient surveys. BMC Med Educ 2017 Aug 21;17(1):136 [FREE Full text] [doi: 10.1186/s12909-017-0967-3] [Medline: 28823250]
33. Chaet D, Clearfield R, Sabin JE, Skimming K, Council on Ethical and Judicial Affairs American Medical Association. Ethical practice in telehealth and telemedicine. J Gen Intern Med 2017 Oct;32(10):1136-1140 [FREE Full text] [doi: 10.1007/s11606-017-4082-2] [Medline: 28653233]
34. The Cleveland Clinic Foundation. Top 10 tips for virtual visits clinician communication. Center for Excellence in Healthcare Communication. 2019. URL: https://my.clevelandclinic.org/-/scassets/files/org/landing/preparing-for-coronavirus/covid-response-digital-health-communication-tips.ashx [accessed 2021-05-22]

Abbreviations

OR: odds ratio
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 28.02.20; peer-reviewed by N Gagliano, J Floch; comments to author 29.06.20; revised version received 15.10.20; accepted 11.03.21; published 21.06.21.

Please cite as:
Rose S, Hurwitz HM, Mercer MB, Hizlan S, Gali K, Yu PC, Franke C, Martinez K, Stanton M, Faiman M, Rasmussen P, Boissy A
Patient Experience in Virtual Visits Hinges on Technology and the Patient-Clinician Relationship: A Large Survey Study With Open-ended Questions
J Med Internet Res 2021;23(6):e18488
URL: https://www.jmir.org/2021/6/e18488
doi:10.2196/18488
PMID:34152276

Original Paper

Determinants of Catalan Public Primary Care Professionals' Intention to Use Digital Clinical Consultations (eConsulta) in the Post–COVID-19 Context: Mixed Methods Study

Francesc Saigí-Rubió[1,2], PhD; Josep Vidal-Alaball[3,4,5], MD, MPH, PhD; Joan Torrent-Sellens[2,6], MSc, PhD; Ana Jiménez-Zarco[2,6], MSc, PhD; Francesc López Segui[7,8], MSc, PhD; Marta Carrasco Hernandez[9], MSc; Xavier Alzaga Reig[10], MD; Josep Maria Bonet Simó[8], MD; Mercedes Abizanda González[10,11], BSc; Jordi Piera-Jimenez[12,13,14], MSc; Oscar Solans[10,12], MD, MSc

[1]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain

[2]Interdisciplinary Research Group on ICTs, Universitat Oberta de Catalunya, Barcelona, Spain

[3]Health Promotion in Rural Areas Research Group, Gerencia Territorial de la Catalunya Central, Institut Catala de la Salut, Sant Fruitos de Bages, Spain

[4]Unitat de Suport a la Recerca de la Catalunya Central, Fundacio Institut Universitari per a la Recerca a l'Atencio Primaria de Salut Jordi Gol i Gurina, Sant Fruitos de Bages, Spain

[5]Faculty of Medicine, University of Vic - Central University of Catalonia, Vic, Spain

[6]Faculty of Economics and Business, Universitat Oberta de Catalunya, Barcelona, Spain

[7]Centre de Recerca en Economia i Salut, Pompeu Fabra University, Barcelona, Spain

[8]Northern Metropolitan Primary Care Directorate, Institut Català de la Salut, Badalona, Spain

[9]Barcelona Primary Care Directorate, Institut Català de la Salut, Barcelona, Spain

[10]Health Department, Catalan Ministry of Health, Barcelona, Spain

[11]Pere Virgili Health Park, Primary Care Management Control, Barcelona, Spain

[12]Digitalization for the Sustainability of the Healthcare System, Sistema de Salut de Catalunya, Barcelona, Spain

[13]Servei Catala de la Salut, Barcelona, Spain

[14]Open Evidence Research Group, Universitat Oberta de Catalunya, Barcelona, Spain

Corresponding Author:
Josep Vidal-Alaball, MD, MPH, PhD
Health Promotion in Rural Areas Research Group
Gerencia Territorial de la Catalunya Central
Institut Catala de la Salut
Carrer Pica d'Estats, 36
Sant Fruitos de Bages, 08272
Spain
Phone: 34 6930040
Email: jvidal.cc.ics@gencat.cat

Abstract

Background: Telemedicine has become a necessary component of clinical practice for the purpose of providing safer patient care during lockdowns due to the COVID-19 pandemic. It has been used to support the health care needs of patients with COVID-19 and routine primary care patients alike. However, this change has not been fully consolidated.

Objective: The objective of this study was to analyze the determinants of health care professionals' intention to use the eConsulta digital clinical consultation tool in the post–COVID-19 context.

Methods: A literature review of the Technology Acceptance Model allowed us to construct a theoretical model and establish a set of hypotheses on the influence of a variety of different factors relating to health care professionals, as well as the institutions where they work, on their intention to use eConsulta. In order to confirm the proposed model, a mixed qualitative and quantitative methodology was used, and a questionnaire was designed to serve as the data collection instrument. The data were analyzed using univariate and bivariate analysis techniques. To confirm the theoretical model, exploratory factor analysis and binary logistic regression were applied.

Results: The most important variables were related to perceived benefits (*B*=2.408) and the type of use that individuals habitually made of eConsulta (*B*=0.715). Environmental pressure (*B*=0.678), experience with technology (*B*=0.542), gender (*B*=0.639), and the degree to which eConsulta had been implemented (*B*=0.266) were other variables influencing the intention to use the tool in the post–COVID-19 context. When replicating the previous analysis according to professional group, experience with technology and gender in the physician group, and experience with tool use and the center where a professional worked in the nurse group, were found to be of considerable importance.

Conclusions: The implementation and use of eConsulta had increased significantly as a consequence of the COVID-19 pandemic, and the majority of health care professionals were satisfied with its use in practice and planned to incorporate it into their practices in the post–COVID-19 context. Perceived benefits and environmental pressure were determining factors in their attitude toward and intention to use eConsulta.

(J Med Internet Res 2021;23(6):e28944) doi:10.2196/28944

KEYWORDS

COVID-19; teleconsultation; eConsultation; eHealth; intention to use; digital health; Technology Acceptance Model; TAM; remote consultation; telemedicine; digital technology; intention; technology assessment; telehealth; pandemic; digital tool

Introduction

Background

Lockdowns and social distancing in response to the high rate of COVID-19 transmission have become the main triggers of a challenging digital transformation in many sectors, especially in health care. In this scenario of extreme crisis, the rapid adoption of digital solutions and technological tools has played an important role in the response to the huge pressure on health care systems [1-3]. Telemedicine has become a necessary component of clinical practice for the purpose of providing safer patient care [4,5], and it has been used to support the health care needs of patients with COVID-19 and routine primary care patients alike [6-10].

While the digital transformation in health care has not been as disruptive as the transformations observed in other industries, the spread of COVID-19 seems to have provided a solid and inevitable reason to fully adopt the digital transformation [11]. However, despite the fact that health care services can largely be provided remotely via digital technologies [12], this change has not become fully consolidated [13,14]. This means that further research contributions are still required in relation to the definition and adoption of new digital care models.

To establish telemedicine in routine health care, acceptance by users—health care professionals—is of vital importance to the effective use of technological resources. A variety of factors may explain why a group adopts a specific technological tool to a lesser or greater extent [15]. Of these, social factors are perceived to be the most complex. Legal limitations, patient indifference, lack of remuneration, and uncoordinated implementation by those responsible for formulating policies are weighty arguments explaining stakeholders' refusal to engage [15,16]. Some studies have noted that these difficulties may be due to a lack of focus in the implementation of such interventions (ie, health care professionals do not see them as either necessary or effective [17]), or the paucity or inconclusive nature of the studies published thus far [18-20].

To understand whether and how the digital technologies adopted to cope with the COVID-19 crisis will continue to be useful in the postemergency phase—beyond research into outcomes thereof (efficiency, care service quality, etc)—it is necessary to understand the determinants of their use. This paper presents an ex-ante analysis and aims to provide evidence on the determinants of use of the Catalan public health care system's eConsulta tool.

eConsulta, which forms part of a personal health folder [21], is an asynchronous teleconsultation tool available to the 7.5 million inhabitants of Catalonia (located in northeastern Spain) and to its primary care professionals. It was launched in 2015 to complement face-to-face care. The tool's implementation has gradually been extended to the entire primary care network (more than 92% of primary care teams [PCTs] have used the tool at some point), and it has recently begun to be introduced into the public hospital setting. However, its rate of use up to March 2020 was low compared to face-to-face consultations (just 0.9% of the total) [22]. Previous studies of telemedicine acceptance by the Catalan public system's health care professionals have suggested that despite being rated positively, especially by nursing staff, the potential technical or organizational disadvantages of the tools were negative predictors of their use [23].

The objective of this study is, therefore, to analyze the determinants of Catalan public primary care professionals' intention to use the eConsulta digital clinical consultation tool in the post–COVID-19 context. These health care professionals (physicians and nurses) work for the Catalan Health Institute (Institut Català de la Salut [ICS]), the main provider of primary care services in Catalonia (providing a 74% coverage of the Catalan population).

Hypotheses and Model

Regarding the theoretical approach, the Technology Acceptance Model (TAM) was used [24,25]. Uptake and use of technological applications in the health care area can be formulated as an acceptance intention, and it is therefore possible to take the TAM approach. In TAMs, technology acceptance is considered a determinant of technology use. Hence, health care professionals' acceptance of a digital technology can be considered a determinant of its use [26-30]. Although TAMs have been widely used to explain the use of many different technologies, health care research has shown them to be highly suited to analyzing eHealth use. Following the TAM methodology, the proposed model contrasts health care professionals' intention to use a digital tool according to two main dimensions: (1) the perceived benefits of the tool's use and (2) its ease of use. Regarding health care professionals' perceived benefits of the tool, the study distinguishes between those connected with improved efficiency in their care activities (better care management or provision, and time savings or reductions [31,32]), and those connected with improved quality in relation to patients [33]. Regarding ease of use, it has been shown that when a health care professional feels that a technological tool is easy to use and does not require any additional training or specific competencies, the intention to use it clearly increases [34]. Within the specific context of the eConsulta digital tool, the first two research hypotheses are:

> *H1. The perceived benefits of eConsulta use have an influence on the intention to use it in the post–COVID-19 context.*
>
> *H2. eConsulta's ease of use has an influence on the intention to use it in the post–COVID-19 context.*

However, despite its generalized use for the purpose of corroborating health care professionals' use of digital tools, the TAM methodology has attracted some criticism, mainly stemming from the fact that it does not take into account the influence of other types of external variable that could increase its explanatory power [35,36]. In this respect, various methodological proposals such as the theory of reasoned action and the theory of planned behavior have noted the appropriateness of considering the external influence exerted by health care professionals' close contacts or environment [37,38]. In particular, it has been found that patients can exert pressure on a professional by asking him or her to use (or not to use) a tool [39]. Likewise, colleagues and other professional groups working in close collaboration may exert a social influence in relation to the tool's use, either because they are users within the same organizational area or because its use has been directly recommended by them [40,41]. Lastly, the organization in which a professional works can have a direct or indirect influence on the tool's use [29]. Thus, the organization itself can encourage its use by establishing policies and offering training, and by increasing the recognition of, or compensating, those professionals who decide to use it:

> *H3. Pressure from other groups (patients, health care professionals, the institution's management team) has an influence on the intention to use eConsulta in the post–COVID-19 context.*

A professional's intention to use the tool will also be affected by professional and demographic variables [42,43]. Thus, an individual's professional profile (physician or nurse) will determine his or her interest in, and use of, the tool. Specifically, an individual's professional profile determines whether the tool is used to care for patients or to carry out management activities. This indirectly means that the type of pressure exerted by the environment to ensure a tool is used, and even the perceived benefits and ease of use of the tool, may be different:

> *H4. The professional profile has an influence on the intention to use eConsulta in the post–COVID-19 context.*

Moreover, the amount of time a professional spends on doing his or her job has an influence on the intention to use the tool. Being older and more professionally experienced may have an influence on the ability or desire to learn about and use technology. Being in a certain occupational category with responsibilities may also have an influence:

> *H5. An individual's experience as a health care professional has an influence on the intention to use eConsulta in the post–COVID-19 context.*
>
> *H6. A health care professional's age has an influence on the intention to use eConsulta in the post–COVID-19 context.*

Similarly, gender has an influence on the intention to use the tool, mainly because there is a gender bias among professionals. In particular, there is a high percentage of women in the nurse group and, among general practitioners, the distribution of men and women varies as the population gets younger, due to the gradual feminization of the medical profession since the end of the 20th century. By the beginning of the 21st century, 70% of all new medical students were women. Since then the figure has risen to 85% [44,45]:

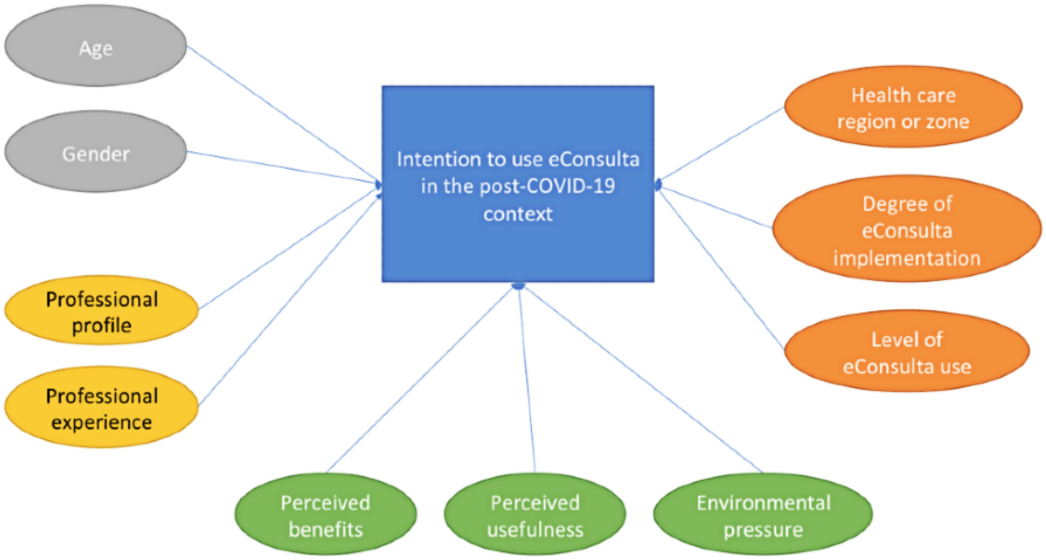
> *H7. A health care professional's gender has an influence on the intention to use eConsulta in the post–COVID-19 context.*

Lastly, the health care center where a professional works also determines the intention to use a tool, mainly for reasons of resource availability and management policies at the center:

> *H8. The health care region or zone in which the health care center where a professional works is located has an influence on the intention to use eConsulta in the post–COVID-19 context.*
>
> *H9. The degree of eConsulta implementation at the health care center where a professional works has an influence on the intention to use it in the post–COVID-19 context.*
>
> *H10. The ease of use of eConsulta (current version) has an influence on the intention to use it in the post–COVID-19 context.*

In summary, we consider that the health care professionals' intention to use the tool in the post–COVID-19 context depends on four main groups of variables: (1) their perception of the tool, (2) external pressure, (3) their profiles, and (4) an additional set of factors linked to the health care center where they work. The proposed model in this regard is shown in Figure 1.

Figure 1. Proposed model of health care professionals' intention to use the eConsulta tool in the post–COVID-19 context.



Methods

Study Design and Sample Selection

As an exploratory study focusing on the analysis of a single health care institution, a mixed qualitative and quantitative methodology was used, and a questionnaire was designed to serve as the data collection instrument (Multimedia Appendix 1). A review of the literature, together with the health care professionals' experience, served as the basis to create the study variables and the metrics used in the first version of the instrument. The measurement instrument was validated following a pretest.

The final questionnaire was organized into four blocks of questions: (1) sociodemographic and professional background, (2) tool use, (3) tool use motivations, and (4) perceived benefits of tool use.

A health care provider distributed the online questionnaire to physicians, nurses, reproductive and sexual health service staff, social workers, and client care staff working in the ICS's various PCTs across Catalonia. Sample selection was random. The questionnaire was sent to all professionals forming the study universe. Participation was voluntary, and no incentives were offered to fill in and return the questionnaire.

A literature review of TAMs allowed us to establish the theoretical model shown in Figure 1. From this, we derived a set of hypotheses about the influence that a variety of different factors relating to both health care professionals and the institutions where they work has on those professionals' intention to use eConsulta. All the variables shown in the hypotheses, which have an influence on the intention to use eConsulta, are described in Table 1.

The data obtained from the sample of health care professionals were analyzed using SPSS Statistics 20 software (IBM Corp). Univariate and bivariate analysis techniques were used. In order to validate the proposed hypotheses and, therefore, to confirm the theoretical model, several multivariate techniques were applied, such as explanatory factor analysis and binary logistic regression.

The study protocol was approved by the University Institute for Primary Care Research Jordi Gol Healthcare Ethics Committee (code 20/026-P).

Table 1. Model variables

Variable	Description
Intention to use eConsulta in the post–COVID-19 context	Dichotomous variable: 0=no, 1=yes
Perceived benefits	Metric variable indicating a professional's degree of perceived benefit from eConsulta use. This variable was created using exploratory factor analysis
Ease of use	Variable metric indicating the degree of eConsulta's ease of use. This variable was created using exploratory factor analysis
Environmental pressure	Variable metric indicating the degree to which a professional perceives that other agents within his or her environment (organization, patients, or colleagues) have an influence on his or her use of eConsulta. This variable was created using exploratory factor analysis
Professional profile	Categorical variable indicating an individual's professional profile, where 1=general practitioner, 2=pediatrician, 3=family nurse, and 4=reproductive and sexual health service nurse
Professional experience	Categorical variable indicating the amount of time a professional has worked in the health care sector, where 1=less than 1 year, 2=2-5 years, 3=6-10 years, 4=11-20 years, and 5=more than 20 years
Gender	Dichotomous variable indicating an individual's gender: 1=man, 2=woman
Age	Categorical variable indicating an individual's age range (in years), where 1=19-29, 2=30-39, 3=40-49, 4=50-59, and 5=60 or above
Health care region or zone in which a professional works	Categorical variable indicating the health care region or zone in which a professional works
Degree of eConsulta implementation	Categorical variable indicating the degree of eConsulta implementation, where 1=not yet fully implemented, 2=fully implemented, and 3=implemented due to COVID-19
Level of eConsulta use	Categorical variable indicating how eConsulta was used, where 1=no use, 2=management use, and 3=management and consultation use

Results

The target sample comprised 18,804 health care professionals working in 285 PCTs. The study included responses from 1189 professionals who had agreed to participate between July 6 and July 31, 2020, the period during which the questionnaire was distributed (margin of error=2.8%; 95%, p=q=0.05).

Sample Profile

Regarding age, we found that the majority of the respondents were over 40 years old. The mean age was 48 years. We found that 33.4% (n=397) were 40-49 years, and 29.2% (n=348) were 50-59 years. It is worth noting that 15.8% (n=188) were over 60 years old, and 16.8% (n=200) were under 30 years old.

The vast majority of the respondents (n=944, 79.2%) were women, meaning that just 20.8% (n=244) were men. In terms of professional profile, 62.2% (n=739) were physicians, with various profiles, and 35.2% (n=418) were other health care professionals—nurses, matrons, or social care workers. Lastly, just 2.2% (n=23) fell into the client care staff category.

In relation to the total, and focusing on the physician group, 43.4% (n=515) were primary care physicians, and 14.8% (n=175) were pediatricians. Regarding the nonphysician health care staff, 26.8% (n=318) of the total were nurses and 7.6% (n=90) were matrons. The majority of these professionals worked in one of three zones of the Barcelona health care region: Metropolitan North (n=250, 21%), Barcelona City (n=218, 18.3%), and Metropolitan South (n=133, 11.2%). Respondents also worked in the health care regions Girona (n=213, 17.9%) and Catalonia Central (n=174, 14.6%).

The majority of the respondents had considerable experience as health care professionals. Indeed, 47.4% (n=563) had over 20 years' experience, and 32.7% (n=390) had 11-20 years' experience. Having gained their experience at several centers, just 48% (n=563) stated that they had spent more than 11 years in the same post (more than 20 years: n=226, 19%; 11-20 years: n=354, 29.7%). Of the remaining respondents, 28% (n=333) had spent 2-5 years in the same post.

Focusing on the workplace, the degree of eConsulta implementation was quite high. Of the total respondents, 44.9% (n=531) indicated that the tool had been fully implemented, and 13.9% (n=164) reported that it had been largely implemented. It is worth noting that 38.6% acknowledged that the pace of the tool's implementation had quickened because of COVID-19.

Regarding the respondents' use of eConsulta, it was found that 60.3% (n=717) had just started using it, and that 26.4% (n=314) used it regularly and intensively. In fact, 40.5% (n=418) acknowledged that they used it at least once a week, and 45.8% (n=473) used it daily. Moreover, the majority of the respondents (n=828, 69.6%) used it to carry out consultations or to appraise test results and make diagnoses, and 16.9% (n=201) acknowledged that they used it for management processes only.

Lastly, it should be noted that the impact of COVID-19 on the tool's use was high. Of the total respondents, 38.6% (n=458) stated that the tool had been implemented in their workplace. On the other hand, 85.7% (n=1018) of respondents agreed or strongly agreed with continuing to use it in the post–COVID-19 context.

Intention to Use eConsulta

In order to identify the factors that have an influence on eConsulta adoption, a principal component exploratory factor analysis was performed. This statistical technique allows the underlying dimensions, constructs, or latent variables of the variables observed in the study to be explored with greater precision. It was on that basis that we endeavored to explain the process of eConsulta adoption.

The values obtained from the statistical tests carried out showed the suitability of the technique employed (Kaiser-Meyer-Olkin index of sampling adequacy=0.928; approximate χ^2=4122.650; Bartlett test of sphericity=171, $P<.001$) (Table 2). Three factors with an eigenvalue higher than 1 were obtained from the analysis, which in total explained 53% of the variance. Likewise, it should be noted that the rotation employed in this case was orthogonal (Varimax) because it was deemed that the different factors might not display any correlation.

Table 2. Rotated component matrix.[a]

Original variables	Perceived usefulness	Environmental pressure	Ease of use (experience with technology)
I feel that eConsulta is useful for managing the calendar with patients, making it easier to schedule visits at times that suit them	0.687	—[b]	—
I feel that it improves the outcomes of my care activities	0.791	—	—
It allows me to offer patients better treatment	0.752	—	—
I talk to other professional colleagues about the benefits of using eConsulta	0.718	—	—
eConsulta has been very useful to me during the COVID-19 pandemic because it has allowed me to care for patients remotely, thus reducing the risk of infection	0.586	—	—
I positively rate the potential benefits that eConsulta use can offer, both for the patients and the PCT[c]/service	0.815	—	—
I promote eConsulta use among my patients	0.744	—	—
As a result of eConsulta implementation in the PCT/service, ways of working have changed, or new ones have been introduced, at individual and group levels	0.632	—	—
The changes made to eConsulta in the COVID-19 pandemic context have made it easier to use	0.536	—	—
I feel that eConsulta is very useful for carrying out my professional activities	—	0.614	—
My colleagues use it often	—	0.662	—
The PCT/service that I work in encourages and facilitates eConsulta use	—	0.514	—
Some of my patients ask me to use it	—	0.541	—
Care professionals can access eConsulta very easily	—	0.629	—
Citizens can access eConsulta very easily	—	0.730	—
I am a habitual user of technology (both professionally and personally)	—	—	0.726
I am a habitual user of social media (both professionally and personally)	—	—	0.798
I have previous experience using telemedicine systems	—	—	0.660
Cronbach alpha	0.916	0.819	0.747
Eigenvalue	7.249	1.534	1.363
Variance explained (%)	28.4	15.3	9.6

[a] Rotation method: Varimax with Kaiser normalization.

[b] Not applicable.

[c] PCT: primary care team.

Three factors emerged from the analysis. The first explained 28.4% of the variance and included the various benefits of eConsulta use, as observed by the professionals. These benefits referred to improvements in the professionals' relationships with patients and in the efficiency of their work. The second factor explained 15.3% of the variance and showed the pressure that third parties—work colleagues, patients, or the institution itself—directly or indirectly exerted on the professionals. Lastly, the third factor was perceived ease of use. This factor explained 9.6% of the variance and showed that some of the professionals who decided to use eConsulta were those who had previous experience of using ICTs (information and communications technologies) (personally) and even telemedicine. A confirmatory factor analysis confirmed the results obtained from the exploratory factor analysis. The results show a goodness of fit (NFI [normed fit index]= 0.897, CFI [comparative fit index]=0.906, TLI [Tucker Lewis index]=0.870, RMSEA [root mean square error of approximation]=0.085).

In order to analyze the influence that these and other factors had on the health care professionals' intention to use eConsulta in the post–COVID-19 context, several logit analyses were performed.

Intention to Use eConsulta in the Post–COVID-19 Context

The various analyses performed showed the model's goodness of fit (Wald=467.731, *P*<.001; Hosmer-Lemeshow=8.525, *P*=.38). Likewise, the model displayed high explanatory power, with a Nagelkerke R^2 value of 0.615.

Regarding the variables in the model, it was found that most of them displayed a direct and significant relationship with the intention to use eConsulta in the post–COVID-19 context (Table 3). The most important variables were those referring to perceived benefits (*B*=2.408) and the type of use that individuals *B*=0.715). Environmental pressure also contributed to the intention to continue using the tool (*B*=0.678), as did experience of using technological tools, some of which were specific to the health care area (*B*=0.542). Lastly, gender had an influence on the intention to use eConsulta (*B*=0.639), as did the degree of its implementation at the health care center where a professional works (*B*=0.266).

Table 3. Logistic regression on eConsulta predictors.

Variable	*B*	SE	Wald	*df*	*P* value	Exp(B)
Age	0.076	0.159	0.227	1	.63	1.079
Gender	0.639	0.308	4.303	1	.04	1.895
Professional profile	–0.280	0.103	7.438	1	.006	0.755
Health care region or zone in which a professional works	0.062	0.048	1.683	1	.20	0.940
Professional experience in the health care area	–0.022	0.164	0.018	1	.89	0.978
Benefits	2.408	–0.199	146.029	1	<.001	11.113
Environmental pressure	0.678	0.129	27.649	1	<.001	1.971
Ease of use	0.542	0.129	17.559	1	<.001	1.720
Degree of eConsulta implementation	0.266	0.149	3.168	1	.08	1.304
Level of eConsulta use	0.715	0.176	16.469	1	<.001	2.045
Constant	0.989	0.894	1.223	1	.27	2.689

The negative value displayed by the profile of the health care professional stands out (*B*=–0.280). This would suggest that the types of activity that different groups carry out within an organizational area had an influence on the rating of, and the intention to use, eConsulta.

With that in mind, and with the aim of confirming or rejecting all the previously proposed hypotheses contained in the model, we replicated the previous analysis for two groups working in the family medicine area: general practitioners and family nurses (there was a total of 507 general practitioners, of whom 89.7% [n=455] used eConsulta and 10.3% [n=52] did not. There were 296 nurses, of whom 87.2% [n=258] used the tool and 12.8% [n=36] did not). Both groups care for adult patients, the majority of whom are advanced in years and are diagnosed with a chronic illness. They are patients whose knowledge and use of ICTs is low.

Considering the different models by group, we found that, among the general practitioner group, all the variables had a direct and positive effect on the intention to use eConsulta. This group was formed by 30.8% (n=156) men and 69.2% (n=351) women, with a mean age (in years) between 40 and 59 years (40-49 years: n=191, 37.7%; 50-59 years: n=151, 29.8%).

Analyzing Table 4, the variable with the greatest weight in the model was perceived benefits (*B*=2.472), followed by gender (*B*=1.011) and the type of use that individuals made of eConsulta (*B*=0.809). Regarding type of use, we found that 92.1% of the physicians made wide use of the tool. Environmental pressure was an important factor in the decision to use it (*B*=0.773). Lastly, an individual's experience of technology (*B*=0.724), together with the degree of eConsulta implementation in his or her workplace (*B*=0.671), also had an influence on the intention to use the tool.

Table 4. Logistic regression on eConsulta predictors (general practitioners).

Variable	*B*	SE	Wald	*df*	*P* value	Exp(B)
Age	0.257	0.310	0.690	1	.41	1.293
Gender	1.011	0.485	4.353	1	.04	2.749
Health care region or zone in which a professional works	0.066	0.096	0.482	1	.49	1.069
Professional experience in the health care area	0.118	0.290	0.166	1	.68	1.126
Benefits	2.472	0.330	56.054	1	<.001	11.850
Environmental pressure	0.773	0.221	12.190	1	<.001	2.166
Ease of use	0.724	0.220	10.864	1	.001	2.062
Degree of eConsulta implementation	0.671	0.332	4.085	1	.04	1.956
Level of eConsulta use	0.809	0.451	3.220	1	.07	2.246
Constant	–2.743	1.958	1.962	1	.16	0.064

The family nurse group (Table 5) was formed by 88.6% (n=285) women, with a mean age between 40 and 59 years (40-49 years: n=104, 33%; 50-59 years: n=95, 30%). Among the variables determining their intention to use eConsulta were, in first place, perceived benefits (*B*=2.100), followed by the use they make of it (*B*=1.362). Here, we found that 50.5% (n=160) stated that they made wide use of the tool, whereas 30% (n=96) only did so for management purposes (Multimedia Appendix 2). Lastly, environmental pressure was a fundamental factor (*B*=0.479). It is worth noting that the health care region or zone in which a health care professional works displayed a value of little importance, which was negative in the model (*B*=–1.91). Of the total cases, 32% (n=101) and 11.8% (n=37) of these professionals worked in the health care zones Metropolitan North and Barcelona City, respectively (Multimedia Appendix 3).

Finally, Table 6 shows the significant variables for the three calculated models.

A comparison of the different models shows that, in all of them, perceived benefits was the variable that had the highest explanatory power. After these, in descending order of importance, were experience of eConsulta use and, lastly, the influence that patients, colleagues, and the institution itself had on the professionals. Regarding these variables, it is worth noting the considerable importance that experience with eConsulta use had for the nurse group (*B*=1.362), compared to the values displayed by this variable in the physician group (*B*=0.809) or, indeed, in the overall model (*B*=0.715).

Meanwhile, some variables were significant for the overall model but were not for the partial models. This was the case for experience with technology, which had a value of *B*=0.724 in the physician group but was not significant in the nurse group. The situation regarding the gender variable was similar; in the physician group, the variable was found to be significant and had a high B value (*B*=1.011), whereas it was not significant in the nurse group.

Table 5. Logistic regression on eConsulta predictors (nurses).

Variable	*B*	SE	Wald	*df*	*P* value	Exp(B)
Age	–0.293	0.316	0.860	1	.35	0.746
Gender	–0.813	0.887	0.840	1	.36	0.444
Health care region or zone in which a professional works	–0.191	0.096	4.001	1	.045	0.826
Professional experience in the health care area	–0.113	0.350	0.105	1	.75	0.893
Benefits	2.100	0.390	29.048	1	<.001	8.169
Environmental pressure	0.479	0.287	2.783	1	.10	1.615
Ease of use	0.161	0.255	0.397	1	.53	1.174
Degree of eConsulta implementation	–0.053	0.309	0.029	1	.87	0.949
Level of eConsulta use	1.362	0.367	13.791	1	<.001	3.906
Constant	3.747	1.909	3.851	1	.05	42.380

Table 6. Significant variables for the three calculated models.

Variable	Overall model		General practitioners		Family nurses	
	B	*P* value	*B*	*P* value	*B*	*P* value
Age	0.076	.63	0.257	.41	–0.293	.35
Gender	0.639	.04	1.011	.04	–0.813	.36
Professional profile	–0.280	.006				
Health care region or zone in which a professional works	0.062	.20	0.066	.49	–0.191	.045
Professional experience in the health care area	–0.022	.89	0.118	.68	–0.113	.75
Benefits	2.408	<.001	2.472	<.001	2.100	<.001
Environmental pressure	0.678	<.001	0.773	<.001	0.479	.10
Ease of use	0.542	<.001	0.724	.001	0.161	.53
Degree of eConsulta implementation	0.266	.08	0.671	.04	–0.053	.87
Level of eConsulta use	0.715	<.001	0.809	.07	1.362	<.001
Constant	0.989	.27	–2.743	.16	3.747	.05

Finally, it should be mentioned that the health care region or zone in which a professional works was a variable that displayed a negative effect on the nurse group's intention to use eConsulta (Table 7).

Table 7. Distribution of the family nurse group by health care region or zone.

Health care region or zone	Nurses, n (%)	Cumulative %
Metropolitan North	45 (15.2)	15.2
Central Catalonia	39 (13.1)	28.3
Barcelona City	56 (18.9)	47.1
Lleida	12 (4.0)	51.2
Metropolitan South	38 (12.8)	64.0
Girona	58 (19.5)	83.5
Camp de Tarragona	24 (8.1)	91.6
Terres de l'Ebre	25 (8.4)	100.0
Total	297 (100.0)	—[a]

[a] Not applicable.

Table 8 shows the hypotheses that were confirmed or rejected for each of the three models.

Table 8. Hypotheses and results.

Hypotheses	Overall model	Physicians	Nurses
H1. The perceived benefits of eConsulta use have an influence on the intention to use it in the post–COVID-19 context.	Yes	Yes	Yes
H2. eConsulta’s ease of use has an influence on the intention to use it in the post–COVID-19 context.	Yes	Yes	No
H3. Pressure from other groups (patients, health care professionals, the institution’s management team) has an influence on the intention to use eConsulta in the post–COVID-19 context.	Yes	Yes	Yes
H4. The professional profile has an influence on the intention to use eConsulta in the post–COVID-19 context.	Yes	—[a]	—
H5. An individual’s experience as a health care professional has an influence on the intention to use eConsulta in the post–COVID-19 context.	No	No	No
H6. A health care professional’s age has an influence on the intention to use eConsulta in the post–COVID-19 context.	Yes	No	No
H7. A health care professional’s gender has an influence on the intention to use eConsulta in the post–COVID-19 context.	Yes	Yes	No
H8. The health care region or zone in which the health care center where a professional works is located has an influence on the intention to use eConsulta in the post–COVID-19 context.	No	No	Yes
H9. The degree of eConsulta implementation at the health care center where a professional works has an influence on the intention to use it in the post–COVID-19 context.	Yes	Yes	No
H10. The ease of use of eConsulta (current version) has an influence on the intention to use it in the post–COVID-19 context.	Yes	Yes	Yes

[a] Not applicable.

Discussion

Principal Findings

The objective of this study was to identify the factors that have an influence on eConsulta adoption, and the influence of these and other factors on the intention to use the tool in the post–COVID-19 context. To that end, a theoretical model based on a modified TAM was used as the analysis tool.

eConsulta has become a key tool for providing remote medical care owing to the challenges posed by the COVID-19 pandemic. The use of this tool has increased significantly since the start of the pandemic, and the majority of health care professionals are now able to consider using it in their routine medical practice, even after the relaxation of social distancing measures and a return to some degree of normality. We focused our research on forecasts for tool use in the post–COVID-19 context because, as soon as social distancing measures are removed, it is likely that professionals will have the option to choose how, that is, by which means, they connect with their patients. This possibility to choose, which is not feasible while social distancing measures are in place, is very significant, not only from the perspective of analyzing the explanatory factors of eConsulta use, but also from the perspective of health policies. With the experience of its mass use during lockdown, we analyzed explanatory factors for the future use of the tool within the context of a greater freedom of choice. The purpose of doing so was to find out whether the lockdowns had changed the factors driving the use of digital tools for the provision of medical care or, conversely, whether the health care professionals perceived such use to be exceptional, with their preferences being more aligned with a return to the prepandemic situation. These results are undoubtedly very useful for the design of public policies on health care delivery via digital technologies.

Our study confirms, as previous studies have done [5], that perceived usefulness was the explanatory factor with the biggest effect on the attitude toward and intention to use eConsulta in the post–COVID-19 context. As the TAM suggests, the significance of this determining factor refers back to the importance of perceived usefulness when the use of a technology needs to be explained [46,47]. Specifically, ICS health care professionals placed importance on improved patient relationships and the efficiency of their work in their intention to use the tool, and on perceived benefits in their intention to use the tool in the post–COVID-19 context. In this respect, several studies have shown that telecare reduces the number of low value-added face-to-face visits, thus providing evidence to support intervention efficiency from the health care provider’s viewpoint [48,49]. It is therefore crucial for the health care system’s best way of operating to be accepted by the professionals working within it.

The next most important variable was the type of use that individuals habitually made of eConsulta. Bearing in mind that the core work of physicians is patient care, whereas that of the nurse group involves either management or carrying out tasks related to communicating with, and sending information to, patients, an increase in perceived usefulness by the latter positively influenced their attitude toward, and increased their intention to use, the tool. This is consistent with many studies on the acceptance of telemedicine solutions by both primary care providers [50] and nursing staff [51].

Perceived ease of use also had a positive impact on the attitude toward using eConsulta, in particular on improving some professionals’ attitude toward and intention to use it because they felt that it would not involve any effort; this was especially so in the physician group [30,52]. As seen in previous studies, general practitioners with prior experience in digital health care technologies were more enthusiastic and optimistic than those who had yet to use it [19,49]. Despite the evidence showing that as knowledge of ICT use increases, the difficulties an individual encounters when using it decrease [53], it is worth noting the little weight that this factor had compared to benefits and environmental pressure. This might be due to either the generalized implementation of this tool by the administration or the rapid digital transformation that this sector has experienced as a consequence of COVID-19. However, experience using technological tools was a variable displaying a direct and significant relationship with the intention to use eConsulta in the post–COVID-19 context. This result highlights the importance of developing staff's competencies for the sustainable adoption of digital solutions in the health care field.

Second in order of importance in the explanation of the tool's use by ICS health care professionals was the pressure that third parties—work colleagues, patients, or the institution itself—directly or indirectly exerted on them. Environmental pressure also contributed to the intention to continue using eConsulta in the post–COVID-19 context. In this respect, we found that when patients had easy access to the tool, it made them ask health care professionals to use it. Access to the personal health folder, a tool that enables citizens to securely access their personal information and online services [54,55], was key. Similarly, the fact that some colleagues rated the tool positively, or actually used it, also had a direct influence on the intention to use it. This could also be attributed to network effects, which are crucial to the adoption of any technology [56]. Lastly, the fact that the health care institution itself had committed to eConsulta implementation was an important reason for adopting it.

Finally, it should be noted that the degree of eConsulta implementation at the health care center where a professional works had an influence on the intention to use it in the post–COVID-19 context, especially among the physician group. This might be due to the fact that the tool's implementation differed at each health care center. Regarding nursing staff, a plausible explanation as to why it did not affect them is that each health care region or zone is independent in terms of the types of activity (patient treatment or management) that the nurse group carries out, so the types of use made of the tool differs. In fact, the nurse group was the most reticent in terms of the continued use of the tool in the postlockdown phase. Research into the implications of eHealth and telemedicine on professional practice has repeatedly shown that implementation of digital practices for the provision of medical care leads to significant changes in the tasks that professionals carry out. For a sustainable eConsulta implementation, the tasks that the physician and nurse groups carry out will undoubtedly need to be reviewed to ensure that the provision of value-added health care is more efficient and of higher quality.

The COVID-19 pandemic has led to the implementation of digital solutions at record speed and with unprecedented impact. The experience that this has provided nurses and physicians with increases the likelihood of them continuing to use it in the post–COVID-19 context. It is worth taking advantage of the impetus that the current crisis has given us to implement at least some of the solutions proposed in the scientific literature. This study's data cannot be extrapolated to other health care systems; however, the results are critical for digital health care policy planners because the success of eConsulta will largely depend on whether health care professionals promote it. In any case, maintaining the drivers of a continued use of digital tools for the provision of medical care must go hand in hand with practices that promote their use by patients. Having professional groups that are active and ready for the digital transformation is of little use if, on the care services demand side, patients continue to opt for face-to-face care as a matter of preference. In this respect, it is important to make further advances in relation to the social dissemination of the strengths of social health care tools, while at the same time putting efforts into reducing their weaknesses, especially the care inequalities that their use may generate.

Limitations

This study has a number of potential limitations. First, it is a survey-based study, subject to the bias response rates that are inherent to all studies based on data of this type. Second, the survey we used was new and unvalidated, utilized to determine eConsulta use by health care professionals. It is our belief, however, that the survey questions posed were of a pragmatic nature, and that the answers faithfully reflected the sentiments of all the groups. Lastly, the survey was administered in less than 1 month in the midst of drastic changes to medical practice brought about by the pandemic, so opinions and preferences may continue to evolve. Notwithstanding the above, we believe that the sample size analyzed and the degree of statistical significance observed together make our results robust.

Conclusions

The implementation and use of eConsulta had increased significantly as a consequence of the COVID-19 pandemic, and the majority of the health care professionals were satisfied with its use in practice and planned to incorporate it into their practices in the post–COVID-19 context. Perceived benefits and environmental pressures were determining factors in the attitude toward and intention to use eConsulta. However, some reticence in terms of the continued use of the tool in the post–COVID-19 context was detected, especially among the nurse group. For this digital transformation in health care to continue beyond the pandemic, it is important to establish connections between health care professionals' use of the tool on one hand and modification of their tasks on the other, and thus improve the quality of their care. In addition, patients must be educated to use the tool more proficiently.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correlation matrix between variables: overall model.
[DOCX File , 22 KB - jmir_v23i6e28944_app1.docx]

Multimedia Appendix 2
Tool use by group.
[DOCX File , 17 KB - jmir_v23i6e28944_app2.docx]

Multimedia Appendix 3
Health care region or zone by group.
[DOCX File , 23 KB - jmir_v23i6e28944_app3.docx]

References

1. Sterling R, LeRouge C. On-Demand Telemedicine as a Disruptive Health Technology: Qualitative Study Exploring Emerging Business Models and Strategies Among Early Adopter Organizations in the United States. J Med Internet Res 2019 Nov 15;21(11):e14304 [FREE Full text] [doi: 10.2196/14304] [Medline: 31730038]
2. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
3. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global Preparedness Against COVID-19: We Must Leverage the Power of Digital Health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
4. Calton B, Abedini N, Fratkin M. Telemedicine in the Time of Coronavirus. J Pain Symptom Manage 2020 Jul;60(1):e12-e14 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.03.019] [Medline: 32240756]
5. Vidal-Alaball J, Acosta-Roja R, Pastor Hernández N, Sanchez Luque U, Morrison D, Narejos Pérez S, et al. Telemedicine in the face of the COVID-19 pandemic. Aten Primaria 2020 Jun;52(6):418-422 [FREE Full text] [doi: 10.1016/j.aprim.2020.04.003] [Medline: 32402477]
6. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
7. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
8. Triana AJ, Gusdorf RE, Shah KP, Horst SN. Technology Literacy as a Barrier to Telehealth During COVID-19. Telemed J E Health 2020 Sep 01;26(9):1118-1119. [doi: 10.1089/tmj.2020.0155] [Medline: 32429770]
9. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Mar 20;26(5):309-313. [doi: 10.1177/1357633x20916567]
10. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the Crisis Into an Opportunity: Digital Health Strategies Deployed During the COVID-19 Outbreak. JMIR Public Health Surveill 2020 May 04;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]
11. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
12. Keesara S, Jonas A, Schulman K. Covid-19 and Health Care's Digital Revolution. N Engl J Med 2020 Jun 04;382(23):e82 [FREE Full text] [doi: 10.1056/nejmp2005835]
13. Herrmann M, Boehme P, Mondritzki T, Ehlers JP, Kavadias S, Truebel H. Digital Transformation and Disruption of the Health Care Sector: Internet-Based Observational Study. J Med Internet Res 2018 Mar 27;20(3):e104 [FREE Full text] [doi: 10.2196/jmir.9498] [Medline: 29588274]
14. Perakslis ED. Strategies for delivering value from digital technology transformation. Nat Rev Drug Discov 2017 Feb;16(2):71-72. [doi: 10.1038/nrd.2016.265] [Medline: 28082744]
15. Choi WS, Park J, Choi JYB, Yang J. Stakeholders' resistance to telemedicine with focus on physicians: Utilizing the Delphi technique. J Telemed Telecare 2019 Jul 01;25(6):378-385. [doi: 10.1177/1357633X18775853] [Medline: 29792080]
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 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]

17. Farr M, Banks J, Edwards HB, Northstone K, Bernard E, Salisbury C, et al. Implementing online consultations in primary care: a mixed-method evaluation extending normalisation process theory through service co-production. BMJ Open 2018 Dec 19;8(3):e019966 [FREE Full text] [doi: 10.1136/bmjopen-2017-019966] [Medline: 29555817]
18. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. The potential of alternatives to face-to-face consultation in general practice, and the impact on different patient groups: a mixed-methods case study. Health Serv Deliv Res 2018 Jun;6(20):1-200. [doi: 10.3310/hsdr06200] [Medline: 29889485]
19. McGrail KM, Ahuja MA, Leaver CA. Virtual Visits and Patient-Centered Care: Results of a Patient Survey and Observational Study. J Med Internet Res 2017 May 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
20. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening. World Health Organization. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1 [accessed 2021-01-17]
21. Saigí F, Cerdá Calafat I, Guanyabens Calvet J, Carrau Vidal E. [Personal health records: the case of the Personal Health Folder of Catalonia (Spain)]. Gac Sanit 2012 Nov;26(6):582-584 [FREE Full text] [doi: 10.1016/j.gaceta.2012.03.005] [Medline: 22554458]
22. Fernández OS, Seguí FL, Vidal-Alaball J, Bonet Simo JM, Vian OH, Cabo PR, et al. Primary Care Doctor Characteristics That Determine the Use of Teleconsultations in the Catalan Public Health System: Retrospective Descriptive Cross-Sectional Study. JMIR Med Inform 2020 Jan 31;8(1):e16484 [FREE Full text] [doi: 10.2196/16484] [Medline: 32012061]
23. Vidal-Alaball J, López Seguí F, Garcia Domingo JL, Flores Mateo G, Sauch Valmaña G, Ruiz-Comellas A, et al. Primary Care Professionals' Acceptance of Medical Record-Based, Store and Forward Provider-to-Provider Telemedicine in Catalonia: Results of a Web-Based Survey. Int J Environ Res Public Health 2020 Jun 08;17(11):4092 [FREE Full text] [doi: 10.3390/ijerph17114092] [Medline: 32521740]
24. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
25. Davis FD. A technology acceptance model for empirically testing new end-user information systems? Theory and results. Massachusetts Institute of Technology. 1985. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2020-11-15]
26. Ducey AJ, Coovert MD. Predicting tablet computer use: An extended Technology Acceptance Model for physicians. Health Policy and Technology 2016 Sep;5(3):268-284. [doi: 10.1016/j.hlpt.2016.03.010]
27. Tao D, Shao F, Wang H, Yan M, Qu X. Integrating usability and social cognitive theories with the technology acceptance model to understand young users' acceptance of a health information portal. Health Informatics J 2020 Jun 11;26(2):1347-1362 [FREE Full text] [doi: 10.1177/1460458219879337] [Medline: 31603378]
28. Lacasta Tintorer D, Flayeh Beneyto S, Manresa JM, Torán-Monserrat P, Jiménez-Zarco A, Torrent-Sellens J, et al. Understanding the discriminant factors that influence the adoption and use of clinical communities of practice: the ECOPIH case. BMC Health Serv Res 2015 Sep 10;15(1):373 [FREE Full text] [doi: 10.1186/s12913-015-1036-4] [Medline: 26358037]
29. Saigi-Rubió F, Jiménez-Zarco A, Torrent-Sellens J. Determinants of the intention to use telemedicine: Evidence from primary care physicians. Int J Technol Assess Health Care 2016 Jul 29;32(1-2):29-36. [doi: 10.1017/s0266462316000015]
30. Saigí-Rubió F, Torrent-Sellens J, Jiménez-Zarco A. Drivers of telemedicine use: comparative evidence from samples of Spanish, Colombian and Bolivian physicians. Implement Sci 2014 Oct 08;9(1):128 [FREE Full text] [doi: 10.1186/s13012-014-0128-6] [Medline: 25293651]
31. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
32. Jennett PA, Affleck Hall L, Hailey D, Ohinmaa A, Anderson C, Thomas R, et al. The socio-economic impact of telehealth: a systematic review. J Telemed Telecare 2003 Dec;9(6):311-320. [doi: 10.1258/135763303771005207] [Medline: 14680514]
33. Kamal SA, Shafiq M, Kakria P. Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). Technology in Society 2020 Feb;60:101212. [doi: 10.1016/j.techsoc.2019.101212]
34. Kissi J, Dai B, Dogbe CS, Banahene J, Ernest O. Predictive factors of physicians' satisfaction with telemedicine services acceptance. Health Informatics J 2020 Sep 19;26(3):1866-1880 [FREE Full text] [doi: 10.1177/1460458219892162] [Medline: 31854222]
35. Bagozzi R. The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift. JAIS 2007 Apr;8(4):244-254. [doi: 10.17705/1jais.00122]
36. Lee SY, Lee K. Factors that influence an individual's intention to adopt a wearable healthcare device: The case of a wearable fitness tracker. Technological Forecasting and Social Change 2018 Apr;129:154-163. [doi: 10.1016/j.techfore.2018.01.002]
37. Schepers J, Wetzels M. A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. Information & Management 2007 Jan;44(1):90-103. [doi: 10.1016/j.im.2006.10.007]
38. Gong Z, Han Z, Li X, Yu C, Reinhardt JD. Factors Influencing the Adoption of Online Health Consultation Services: The Role of Subjective Norm, Trust, Perceived Benefit, and Offline Habit. Front Public Health 2019 Oct 4;7:286 [FREE Full text] [doi: 10.3389/fpubh.2019.00286] [Medline: 31637229]
39. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and Facilitators That Influence Telemedicine-Based, Real-Time, Online Consultation at Patients' Homes: Systematic Literature Review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]

40. Erickson SM, Rockwern B, Koltov M, McLean RM, Medical Practice Quality Committee of the American College of Physicians. Putting Patients First by Reducing Administrative Tasks in Health Care: A Position Paper of the American College of Physicians. Ann Intern Med 2017 May 02;166(9):659-661 [FREE Full text] [doi: 10.7326/M16-2697] [Medline: 28346948]
41. Liu S, Reese TJ, Kawamoto K, Del Fiol G, Weir C. A systematic review of theoretical constructs in CDS literature. BMC Med Inform Decis Mak 2021 Mar 17;21(1):102 [FREE Full text] [doi: 10.1186/s12911-021-01465-2] [Medline: 33731089]
42. Zhang X, Han X, Dang Y, Meng F, Guo X, Lin J. User acceptance of mobile health services from users' perspectives: The role of self-efficacy and response-efficacy in technology acceptance. Inform Health Soc Care 2017 Mar 26;42(2):194-206. [doi: 10.1080/17538157.2016.1200053] [Medline: 27564428]
43. Taylor D, Bury N, Campling N, Carter S, Garfied S, Newbould J, et al. A Review of the use of the Health Belief Model (HBM), the Theory of Reasoned Action (TRA), the Theory of Planned Behaviour (TPB) and the Trans-Theoretical Model (TTM) to study and predict health related behaviour change. Department of Health, National Institute for Health and Clinical Excellence. 2007 Feb. URL: https://warwick.ac.uk/fac/sci/med/study/ugr/mbchb/phase1_08/semester2/healthpsychology/nice-doh_draft_review_of_health_behaviour_theories.pdf [accessed 2020-11-15]
44. Arrizabalaga P, Bruguera M. [The feminization and the profession of Medicine]. Med Clin (Barc) 2009 Jul 04;133(5):184-186. [doi: 10.1016/j.medcli.2009.01.027] [Medline: 19426998]
45. Paik JE. The Feminization of Medicine. JAMA 2000 Feb 02;283(5):666. [doi: 10.1001/jama.283.5.666-jms0202-2-1]
46. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2014 Feb 16;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
47. Saigi-Rubió F, Jiménez-Zarco A, Torrent-Sellens J. Determinants of the intention to use telemedicine: Evidence from primary care physicians. Int J Technol Assess Health Care 2016 Jul 29;32(1-2):29-36. [doi: 10.1017/s0266462316000015]
48. López Seguí F, Walsh S, Solans O, Adroher Mas C, Ferraro G, García-Altés A, et al. Teleconsultation Between Patients and Health Care Professionals in the Catalan Primary Care Service: Message Annotation Analysis in a Retrospective Cross-Sectional Study. J Med Internet Res 2020 Sep 17;22(9):e19149 [FREE Full text] [doi: 10.2196/19149] [Medline: 32687477]
49. López Seguí F, Vidal-Alaball J, Sagarra Castro M, García-Altés A, García Cuyàs F. General Practitioners' Perceptions of Whether Teleconsultations Reduce the Number of Face-to-face Visits in the Catalan Public Primary Care System: Retrospective Cross-Sectional Study. J Med Internet Res 2020 Mar 16;22(3):e14478 [FREE Full text] [doi: 10.2196/14478] [Medline: 32175914]
50. Tudiver F, Wolff LT, Morin PC, Teresi J, Palmas W, Starren J, et al. Primary care providers' perceptions of home diabetes telemedicine care in the IDEATel project. J Rural Health 2007 Jan;23(1):55-61. [doi: 10.1111/j.1748-0361.2006.00068.x] [Medline: 17300479]
51. Mullen-Fortino M, DiMartino J, Entrikin L, Mulliner S, Hanson CW, Kahn JM. Bedside nurses' perceptions of intensive care unit telemedicine. Am J Crit Care 2012 Jan 31;21(1):24-31; quiz 32. [doi: 10.4037/ajcc2012801] [Medline: 22210696]
52. Idoga PE, Toycan M, Nadiri H, Çelebi E. Assessing factors militating against the acceptance and successful implementation of a cloud based health center from the healthcare professionals' perspective: a survey of hospitals in Benue state, northcentral Nigeria. BMC Med Inform Decis Mak 2019 Feb 19;19(1):34 [FREE Full text] [doi: 10.1186/s12911-019-0751-x] [Medline: 30782166]
53. Kalayou MH, Endehabtu BF, Tilahun B. The Applicability of the Modified Technology Acceptance Model (TAM) on the Sustainable Adoption of eHealth Systems in Resource-Limited Settings. JMDH 2020 Dec;Volume 13:1827-1837. [doi: 10.2147/jmdh.s284973]
54. Model d'Atenció No Presencial en el Sistema Sanitari de Catalunya. Departament de Salut, Barcelona. 2014. URL: http://salutweb.gencat.cat/web/.content/_ambits-actuacio/Linies-dactuacio/model_assistencial/MANP2013_2016.pdf [accessed 2021-06-09]
55. La Meva Salut (Personal Health Folder). URL: https://lamevasalut.gencat.cat/ [accessed 2021-06-09]
56. Torrent-Sellens J. Knowledge Products and Network Externalities: Implications for the Business Strategy. J Knowl Econ 2012 Nov 8;6(1):138-156. [doi: 10.1007/s13132-012-0122-7]

Abbreviations

CFI: comparative fit index
ICS: Institut Català de la Salut
ICT: information and communications technology
NFI: normed fit index
PCT: primary care team
RMSEA: root mean square error of approximation
TAM: Technology Acceptance Model
TLI: Tucker Lewis index

Edited by C Basch; submitted 19.03.21; peer-reviewed by E van der Velde, S Liu; comments to author 14.04.21; revised version received 18.05.21; accepted 31.05.21; published 24.06.21.

Please cite as:
Saigí-Rubió F, Vidal-Alaball J, Torrent-Sellens J, Jiménez-Zarco A, López Segui F, Carrasco Hernandez M, Alzaga Reig X, Bonet Simó JM, Abizanda González M, Piera-Jimenez J, Solans O
Determinants of Catalan Public Primary Care Professionals' Intention to Use Digital Clinical Consultations (eConsulta) in the Post–COVID-19 Context: Mixed Methods Study
J Med Internet Res 2021;23(6):e28944
URL: https://www.jmir.org/2021/6/e28944/
doi:10.2196/28944
PMID:

Original Paper

Association of Smartphone Ownership and Internet Use With Markers of Health Literacy and Access: Cross-sectional Survey Study of Perspectives From Project PLACE (Population Level Approaches to Cancer Elimination)

Sachiko M Oshima[1*], BA; Sarah D Tait[1*], BA; Samantha M Thomas[2,3], MS; Oluwadamilola M Fayanju[3,4], MA, MD, MPHS; Kearston Ingraham[3], MPH; Nadine J Barrett[3,5,6], MA, MS, PhD; E Shelley Hwang[3,4], MPH, MD

[1]Duke University School of Medicine, Durham, NC, United States
[2]Department of Biostatistics & Bioinformatics, Duke University School of Medicine, Durham, NC, United States
[3]Duke Cancer Institute, Durham, NC, United States
[4]Department of Surgery, Duke University School of Medicine, Durham, NC, United States
[5]Department of Family Medicine and Community Health, Duke University School of Medicine, Durham, NC, United States
[6]Duke Clinical Translation Science Institute, Duke University School of Medicine, Durham, NC, United States
[*]these authors contributed equally

Corresponding Author:
E Shelley Hwang, MPH, MD
Department of Surgery
Duke University School of Medicine
465 Seeley Mudd Building
Durham, NC, 27710
United States
Phone: 1 919 684 6849
Email: shelley.hwang@duke.edu

Abstract

Background: Telehealth is an increasingly important component of health care delivery in response to the COVID-19 pandemic. However, well-documented disparities persist in the use of digital technologies.

Objective: This study aims to describe smartphone and internet use within a diverse sample, to assess the association of smartphone and internet use with markers of health literacy and health access, and to identify the mediating factors in these relationships.

Methods: Surveys were distributed to a targeted sample designed to oversample historically underserved communities from April 2017 to December 2017. Multivariate logistic regression was used to estimate the association of internet and smartphone use with outcomes describing health care access and markers of health literacy for the total cohort and after stratifying by personal history of cancer. Health care access was captured using multiple variables, including the ability to obtain medical care when needed. Markers of health literacy included self-reported confidence in obtaining health information.

Results: Of the 2149 participants, 1319 (61.38%) were women, 655 (30.48%) were non-Hispanic White, and 666 (30.99%) were non-Hispanic Black. The median age was 51 years (IQR 38-65). Most respondents reported using the internet (1921/2149, 89.39%) and owning a smartphone (1800/2149, 83.76%). Compared with the respondents with smartphone or internet access, those without smartphone or internet access were more likely to report that a doctor was their most recent source of health information (344/1800, 19.11% vs 116/349, 33.2% for smartphone and 380/1921, 19.78% vs 80/228, 35.1% for internet, respectively; both *P*<.001). Internet use was associated with having looked for information on health topics from any source (odds ratio [OR] 3.81, 95% CI 2.53-5.75) and confidence in obtaining health information when needed (OR 1.83, 95% CI 1.00-3.34) compared with noninternet users. Smartphone owners had lower odds of being unable to obtain needed medical care (OR 0.62, 95% CI 0.40-0.95) than nonsmartphone owners. Among participants with a prior history of cancer, smartphone ownership was significantly associated with higher odds of confidence in ability to obtain needed health information (OR 5.63, 95% CI 1.05-30.23) and lower odds of inability to obtain needed medical care (OR 0.17, 95% CI 0.06-0.47), although these associations were not significant among participants without a prior history of cancer.

Conclusions: We describe widespread use of digital technologies in a community-based cohort, although disparities persist. In this cohort, smartphone ownership was significantly associated with ability to obtain needed medical care, suggesting that the use of smartphone technology may play a role in increasing health care access. Similarly, major illnesses such as cancer have the potential to amplify health engagement. Finally, special emphasis must be placed on reaching patient populations with limited digital access, so these patients are not further disadvantaged in the new age of telehealth.

(J Med Internet Res 2021;23(6):e24947) doi:10.2196/24947

KEYWORDS

telehealth; technology; health literacy; access to health care; mobile phone

Introduction

Background

The COVID-19 pandemic has highlighted the increasing dependence of the health system on telemedicine because providers have relied on telehealth to provide patient care while minimizing the risk of viral transmission [1-3]. Telehealth, or the use of technology—including the internet and mobile phones—to enable or improve health or health care [4,5], has expanded in scope and capabilities in recent decades. Patients can now access personal health information through patient portals [6], look up health information independently without professional medical guidance [7-11], and provide feedback to other consumers on their experiences with certain hospitals or providers through these media [12]. Studies have shown that the use of technology in the health sector can improve intervention efficacy, patient satisfaction, and, ultimately, clinical outcomes [13-17]. Technology has assumed a growing role within the health care landscape, with more than 60% of all US health care institutions using at least one form of telehealth [18], with some engaging in more telehealth visits than in-person visits [19].

Although telehealth has shown promise as a means to expand access to care [20,21], documented disparities persist among individuals who engage with health technology. The gap in access to technology based on social, physical, and societal factors is often referred to as the *digital divide* [22]. Previous work has characterized this divide extensively, with older, less-educated individuals having lower use of internet, mobile phone, and smartphone technologies [23-25] and preferring to receive health information through printed media compared with younger and more-educated individuals [26]. Furthermore, bridging the divide involves not only addressing gaps in physical access to technology, known as the *first digital divide*, but also the reliability of access and technological literacy, known as the *second digital divide* [27-30].

The second digital divide reflects the well-documented association between lower rates of phone and internet use with decreased health literacy [31-33]. This relationship seems to be multifaceted, with recent studies demonstrating an association not only between health literacy and the likelihood of technology adoption, but also between health literacy and the use of health information technology [34,35]. These findings allude to the evolving body of literature characterizing the relationship between digital technology and health outcomes. Preliminary studies have demonstrated that increased internet use can positively affect perceived health outcomes as well as the use of health care services. These differences in the outcomes, benefits, and impacts of technology use represent the *third digital divide* [30,36]. Work thus far describing the relationship between health literacy and health outcomes has led to mixed, inconclusive results [37].

A better understanding of the nature and ramifications of the digital divide is critical because patient populations who are already disadvantaged are at increased risk of being marginalized as the health system evolves to increasingly rely on telehealth to deliver care. Studies have already been published describing the widening inequity resulting from the health system's increased reliance on telehealth as a result of the COVID-19 pandemic and the negative repercussions that most frequently and meaningfully affect vulnerable communities [38]. Furthermore, given the United States' ongoing poor performance in health care access and affordability compared with peer developed nations [39], the US health care system is particularly vulnerable to widening disparities as a result of the stress of dealing with the COVID-19 pandemic.

Objective

This study seeks to provide valuable data from a racially diverse cohort of North Carolina residents to further characterize the current state of digital technology use and to explore the relationship among technological access, health literacy, and health care access. Furthermore, this study builds on preliminary research assessing how major illnesses such as cancer relate to internet- and smartphone-use behaviors and their associations with markers of health literacy and health care access [40,41].

Methods

Study Design and Participants

This cross-sectional study used survey data derived from a community health assessment initiative, Project PLACE (Population Level Approaches to Cancer Elimination). The survey was administered to a targeted convenience sample of diverse populations across a predefined patient catchment area in proximity to a National Cancer Institute–designated comprehensive cancer center in North Carolina, and it aimed to oversample historically underserved communities [42]. Data were collected from April 2017 to December 2017. The study protocol was approved by the Duke University Institutional Review Board (#00062661).

Most of the participants were recruited from community organizations located in Durham, Wake, Vance, Alamance, and Johnston counties in central North Carolina. Community navigators worked in conjunction with 24 community partners to collect survey data using a multimodal approach. The community partners—comprising community organizations, faith organizations, community outreach programs, and a health clinic—distributed the surveys to their constituents at 47 different community events. Community partners received stipends (US $10 per survey, up to US $2000) for their collaboration. Survey participants were offered items valued at US $5 or less (eg, water bottle and tote bag) for their participation [43].

Survey Design

The 91-item, self-administered survey was available in English, Spanish, and Chinese and could be completed on paper or on the web. The survey items included a combination of program-specific and pre-existing validated measures sourced from national surveys, including the Health Information National Trend Survey and the National Health Interview Study [44]. Data from select survey items that assessed sociodemographic factors, personal cancer history, patterns of mobile phone and internet use, and markers of health literacy and health care access were used in this study (Multimedia Appendix 1). The place of residence, metropolitan or nonmetropolitan, was captured with self-reported ZIP code that was coded using Rural-Urban Continuum Codes (RUCC) in which RUCC 1 to 3 were coded as metropolitan and RUCC 4 to 9 were coded as nonmetropolitan [45,46].

Independent Variables

The primary independent variables included internet use and smartphone ownership. Internet use was captured with this question: “Do you ever go on-line to access the Internet or World Wide Web, or to send and receive e-mail?” Smartphone ownership was captured with this question: “Do you currently have a Smart phone such as an iPhone, Android, Blackberry or Windows phone?”

Outcome Measures

The primary outcome measures were health care access and markers of health literacy. Health care access was captured with several items, including having a usual place of care when sick or in need of advice regarding health, the type of place attended for care, being unable to get care when needed, health insurance status, and participation in medical research. Markers of health literacy were captured with several additional items, including having looked for information on health topics, the source used for information on health topics, self-reported confidence in the ability to obtain health information, and self-reported understanding of numerical information (ie, numeracy). The variable “self-reported numeracy” was dichotomized by grouping survey responses 1-3 as “low self-reported numeracy” and survey responses 4-6 as “high self-reported numeracy.” The variable “confidence in ability to obtain health information if needed” was dichotomized with the survey responses “completely confident,” “very confident,” and “somewhat confident” grouped together compared with the grouping of “a little confident” and “not at all confident” (Multimedia Appendix 1).

Statistical Analysis

Descriptive statistics were used to summarize the study sample. Continuous and categorical variables were summarized as median (IQR), where IQR is reported as first quartile value–third quartile value, and n (%), respectively, by smartphone ownership and internet use. Differences were tested using the chi-square test or Fisher exact test for categorical variables, as appropriate, and the two-tailed *t* test for continuous variables.

Logistic regression was used to estimate the association of smartphone ownership and internet use, respectively, with health care use and literacy variables after adjustment for sociodemographic factors both in the total study cohort and after stratifying by personal history of cancer. Covariates were selected based on univariate analysis ($P<.10$). Only respondents with complete data were included in each analysis, and effective sample sizes are indicated for each table and figure. No adjustments were made for multiple comparisons. Two-tailed tests were used for all analyses, and the threshold for significance was set at $P<.05$. All statistical analyses were conducted using SAS version 9.4 (SAS Institute).

Results

Participant Sociodemographic and Digital Technology Use Characteristics

A total of 2315 surveys were completed. Of these 2315 surveys, 2149 (92.83%) respondents answered all 3 questions pertaining to mobile phone ownership, smartphone ownership, and internet use. Demographic, personal health history, and smartphone and internet use characteristics are shown in Table 1.

Table 1. Baseline characteristics of study cohort (N=2149)[a].

Characteristics	All respondents	Smartphone ownership				Internet use			
		No (n=349), n (%)	Yes (n=1800), n (%)	*P* value	Chi-square (*df*)	No (n=228), n (%)	Yes (n=1921), n (%)	*P* value	Chi-square (*df*)
Age, median (IQR)	51 (38-65)	68 (58-76)	48 (36-61)	<.001	272.3 (1)	67 (55-76)	49 (37-63)	<.001	130.1 (1)
Gender, n (%)				.32	N/A[b]			.99[c]	N/A
Female	1319 (61.38)	216 (16.4)	1103 (83.62)			133 (10.1)	1186 (89.92)		
Male	732 (34.06)	106 (14.5)	626 (85.52)			74 (10.1)	658 (89.89)		
Other	8 (0.37)	0 (0)	8 (100)			0 (0)	8 (100)		
Race and ethnicity, n (%)				.001	19.1 (4)			.94	0.8 (4)
Hispanic	300 (13.96)	52 (17.3)	248 (82.67)			29 (9.7)	271 (90.33)		
Non-Hispanic Asian	202 (9.4)	17 (8.4)	185 (91.58)			19 (9.4)	183 (90.59)		
Non-Hispanic Black	666 (30.99)	78 (11.7)	588 (88.29)			64 (9.6)	602 (90.39)		
Non-Hispanic White	655 (30.48)	118 (18)	537 (81.98)			55 (8.4)	600 (91.6)		
Other	82 (3.82)	10 (12.2)	72 (87.8)			8 (9.8)	74 (90.24)		
Income adequacy, n (%)				.02	5.8 (1)			.27	1.2 (1)
Living comfortably or getting by on present income	1629 (75.8)	234 (14.4)	1395 (85.64)			156 (9.6)	1473 (90.42)		
Finding it difficult or very difficult on present income	366 (17.03)	71 (19.4)	295 (80.6)			42 (11.5)	324 (88.52)		
Education level, n (%)				<.001	138.1 (2)			<.001	140.6 (2)
High school or less	487 (22.66)	151 (31)	336 (68.99)			114 (23.4)	373 (76.59)		
Post high school training or some college	535 (24.9)	84 (15.7)	451 (84.3)			48 (8.97)	487 (91.03)		
College graduate or higher	1017 (47.32)	78 (7.7)	939 (92.33)			40 (3.9)	977 (96.07)		
Occupational status, n (%)				<.001	251.5 (4)			<.001	100.6 (4)
Disabled	83 (3.86)	30 (36.1)	53 (63.86)			19 (22.9)	64 (77.11)		
Employed	1192 (55.47)	77 (6.5)	1115 (93.54)			66 (5.5)	1126 (94.46)		
Unemployed	89 (4.14)	18 (20.2)	71 (79.78)			5 (5.6)	84 (94.38)		
Retired	456 (21.22)	163 (35.7)	293 (64.25)			93 (20.4)	363 (79.61)		
Other	190 (8.84)	20 (10.5)	170 (89.47)			15 (7.9)	175 (92.11)		
Insurance status, n (%)				<.001	196.3 (3)			<.001	70.7 (3)
Private	1048 (48.77)	56 (5.3)	992 (94.66)			51 (4.9)	997 (95.13)		
Public	626 (29.13)	192 (30.7)	434 (69.33)			108 (17.3)	518 (82.75)		
Insured, unknown type	95 (4.42)	18 (18.9)	77 (81.05)			14 (14.7)	81 (85.26)		
None	266 (12.38)	44 (16.5)	222 (83.46)			27 (10.2)	239 (89.85)		
Location type, n (%)				<.001	39.1 (1)			<.001	24.4 (1)
Metropolitan	1696 (78.92)	222 (13.1)	1474 (86.91)			143 (8.4)	1553 (91.57)		
Nonmetropolitan	373 (17.36)	97 (26)	276 (73.99)			63 (16.9)	310 (83.11)		
Personal history of cancer, n (%)				<.001	35.8 (1)			.07	3.3 (1)
No	1761 (81.95)	247 (14)	1514 (85.97)			175 (9.9)	1586 (90.06)		

Characteristics	All respondents	Smartphone ownership				Internet use			
		No (n=349), n (%)	Yes (n=1800), n (%)	*P* value	Chi-square (*df*)	No (n=228), n (%)	Yes (n=1921), n (%)	*P* value	Chi-square (*df*)
Yes	340 (15.82)	92 (27.1)	248 (72.94)			45 (13.2)	295 (86.76)		

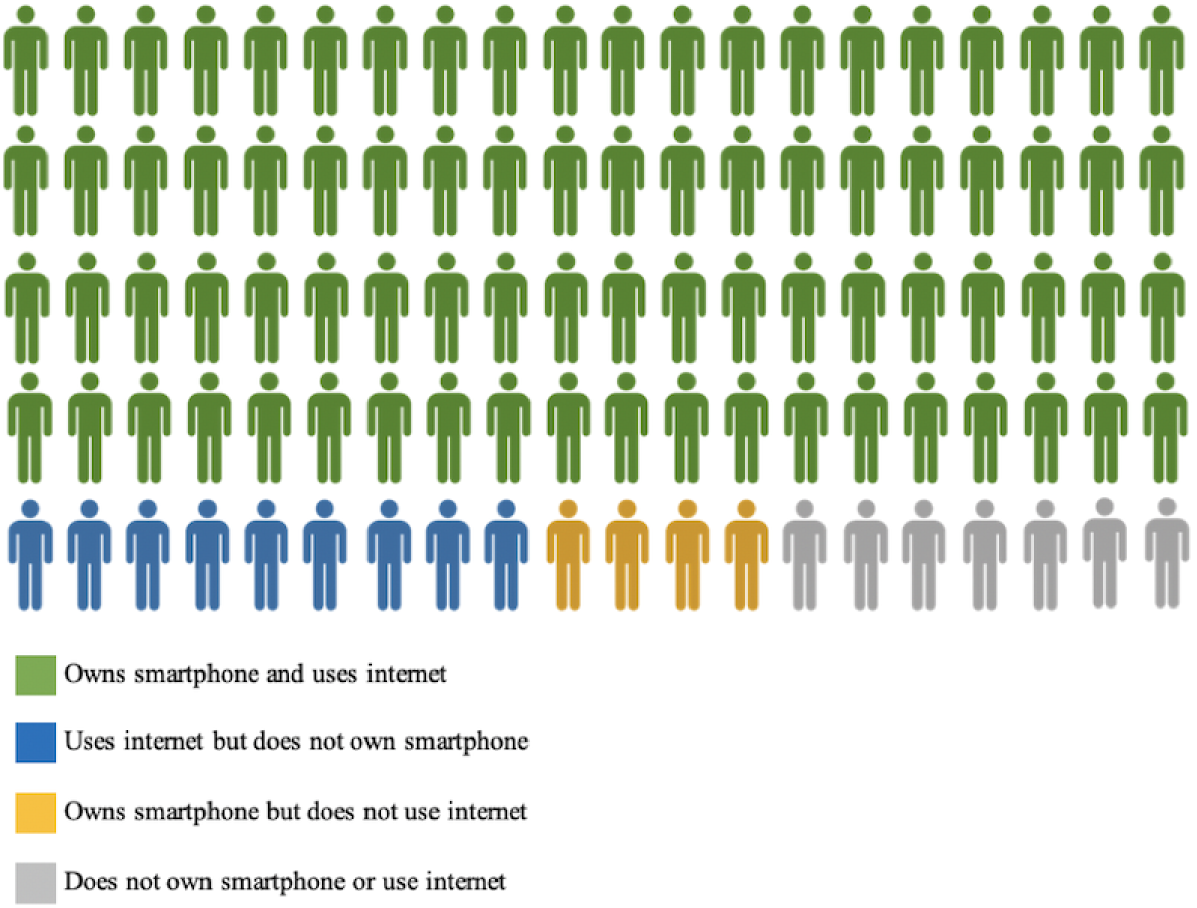
[a] Test statistics and *df* are presented for the chi-square and *t* test *P* values only.

[b] N/A: not applicable.

[c] Fisher exact test *P* value.

The median age was 51 years (IQR 38-65), and most of the participants were women (1319/2149, 61.38%). The racial and ethnic distribution was 13.96% (300/2149) Hispanic, 9.4% (202/2149) non-Hispanic Asian, 30.99% (666/2149) non-Hispanic Black, 30.48% (655/2149) non-Hispanic White, and 3.82% (82/2149) other, whereas 11.35% (244/2149) chose not to respond to this question. Overall, 75.8% (1629/2149) of the participants reported living comfortably or getting by on their present income, and 47.32% (1017/2149) reported an education level of college graduate or higher. Most of the participants lived in a metropolitan area (1696/2149, 78.92%), were employed (1192/2149, 55.47%), and did not have a history of cancer (1761/2149, 81.95%). In total, 89.39% (1921/2149) of the respondents reported using the internet, 96.32% (2070/2149) reported owning a mobile phone, and 83.76% (1800/2149) reported owning a smartphone (Figure 1).

Figure 1. Prevalence of smartphone ownership and internet use among study cohort.



Sociodemographic Factors Associated With Digital Access

Specific characteristics distinguished smartphone owners and internet users from nonsmartphone owners and noninternet users (Table 1). Compared with smartphone owners and internet users, the respondents who reported not owning a smartphone or not using the internet were more likely to be older (both *P*<.001), less-educated (both *P*<.001), disabled (both *P*<.001), retired (both *P*<.001), publicly insured (both *P*<.001), and living in a nonmetropolitan area (both *P*<.001). The respondents who reported not owning a smartphone were also more likely to find it difficult or very difficult to get by on current income compared with smartphone owners (*P*=.02).

Digital Technology Use and Markers of Health Literacy

Smartphone ownership and internet use demonstrated significant thematic associations with markers of health literacy (Table 2).

Table 2. Markers of health literacy and health access by smartphone ownership and internet use (N=2149).[a]

Survey response	All respondents, n (%)	Smartphone ownership				Internet use			
		No (n=349), n (%)	Yes (n=1800), n (%)	*P* value	Chi-square (*df*)	No (n=228), n (%)	Yes (n=1921), n (%)	*P* value	Chi-square (*df*)
Has looked for information on health or medical topics from any source				.006	7.7 (1)			<.001	51.8 (1)
No	249 (11.59)	55 (22.1)	194 (77.91)			59 (23.7)	190 (76.31)		
Yes	1867 (86.88)	284 (15.2)	1583 (84.79)			164 (8.8)	1703 (91.22)		
Is confident in ability to get health information if needed				.001	10.9 (1)			.003	8.7 (1)
Completely, very, or somewhat confident	1697 (78.97)	243 (14.3)	1454 (85.68)			139 (8.2)	1558 (91.81)		
A little or not at all confident	142 (6.61)	35 (24.6)	107 (75.35)			22 (15.5)	120 (84.51)		
Self-reported numeracy (1 is low, 6 is high)				<.001	18.7 (1)			<.001	32.9 (1)
1-3	710 (33.04)	146 (20.6)	564 (79.44)			110 (15.5)	600 (84.51)		
4-6	1378 (64.12)	183 (13.3)	1195 (86.72)			103 (7.5)	1275 (92.53)		
Has a usual place for health care or advice				.37	1.9 (2)			.47	1.5 (2)
Yes	1671 (77.76)	268 (16)	1403 (83.96)			173 (10.4)	1498 (89.65)		
There is more than one place	164 (7.63)	26 (15.9)	138 (84.15)			14 (8.5)	150 (91.46)		
There is no place	211 (9.82)	26 (12.3)	185 (87.68)			17 (8.1)	194 (91.94)		
Usual place for health care				.18	4.8 (3)			.35	3.2 (3)
Hospital emergency room	63 (2.93)	14 (22.2)	49 (77.78)			8 (12.7)	55 (87.3)		
Hospital outpatient department, clinic, or health center; doctor's office or HMO[b]	1762 (81.99)	280 (15.9)	1482 (84.11)			178 (10.1)	1584 (89.9)		
There is no one place	99 (4.61)	11 (11.1)	88 (88.89)			8 (8.1)	91 (91.92)		
Some other place	40 (1.86)	9 (22.5)	31 (77.5)			7 (17.5)	33 (82.5)		
Needed medical care but could not get in within the last 12 months				<.001	17.8 (1)			.41	0.7 (1)
No	1721 (80.08)	242 (14.1)	1479 (85.94)			165 (9.6)	1556 (90.41)		
Yes	306 (14.24)	72 (23.5)	234 (76.47)			34 (11.1)	272 (88.89)		
Asked to participate in a clinical trial or medical research				.76	0.0 (1)			.002	9.2 (1)
No	1599 (74.41)	255 (15.9)	1344 (84.05)			183 (11.4)	1416 (88.56)		
Yes	494 (22.99)	76 (15.4)	418 (84.62)			33 (6.7)	461 (93.32)		

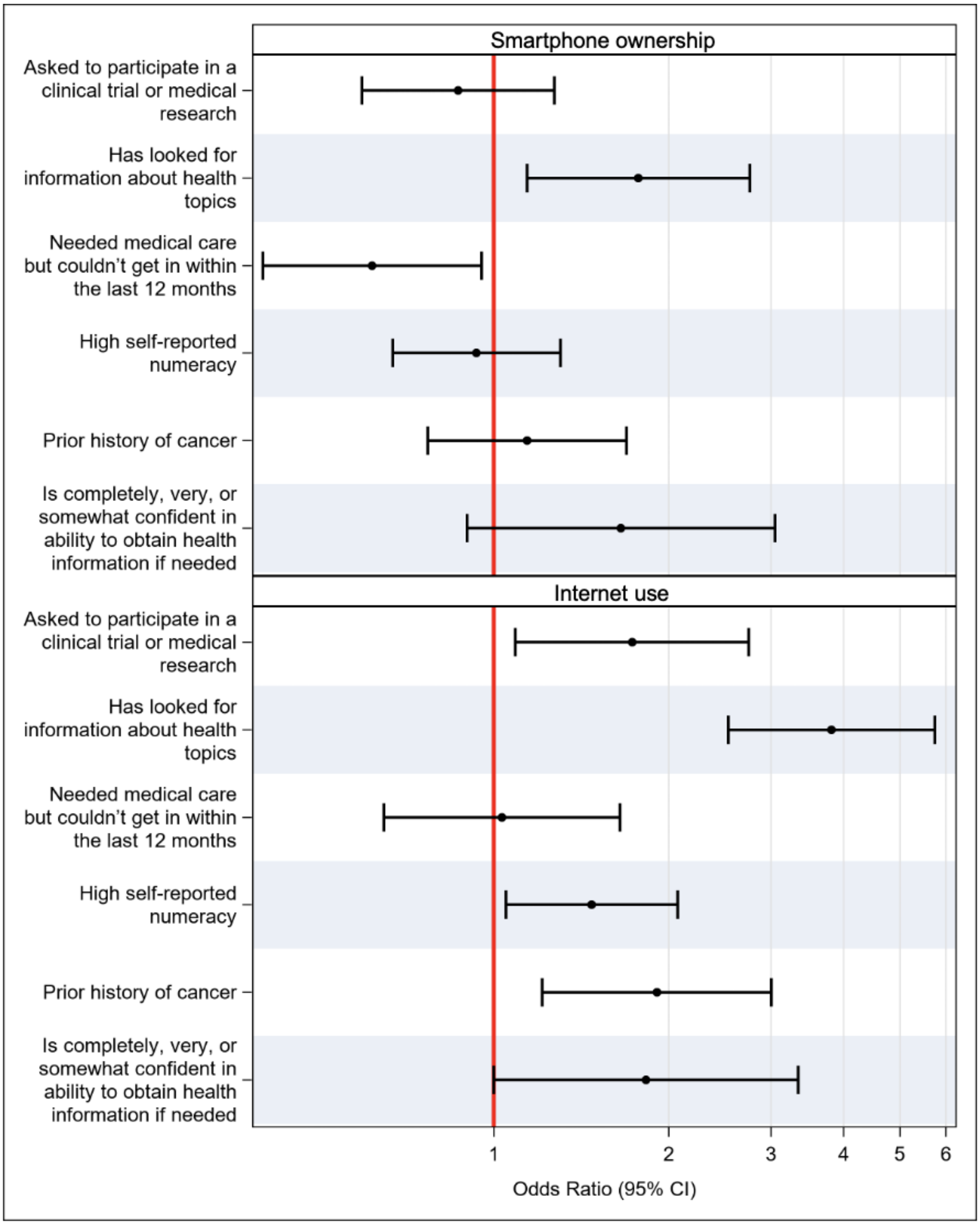
[a] Percentages may not add up to 100 owing to rounding or missing values.

[b] HMO: health maintenance organization.

On univariate analysis, both smartphone ownership and internet use were associated with higher self-reported numeracy (both *P*<.001), confidence in obtaining health information if needed (*P*=.001 and *P*=.003, respectively), and having looked for information on health topics from any source (*P*=.006 and *P*<.001, respectively).

In the adjusted analysis, smartphone ownership was associated with higher odds of having looked for health information (odds ratio [OR] 1.77, 95% CI 1.14-2.76), when controlling for age, race and ethnicity, nativity, language spoken at home, income adequacy, occupational status, education level, insurance status, and rurality (Figure 2). Similarly, internet use was associated with higher odds of having looked for information on health topics (OR 3.81, 95% CI 2.53-5.75), confidence in obtaining health information if needed (OR 1.83, 95% CI 1.00-3.34), and self-reported numeracy (OR 1.47, 95% CI 1.05-2.07) when controlling for age, occupational status, education level, insurance status, and rurality (Figure 2). Smartphone ownership models were adjusted for the following covariates: age, race and ethnicity, nativity, spoken language, income adequacy, occupational status, education level, insurance status, and rurality. Covariates were selected based on univariate analysis (*P*<.10 on univariate analysis). Internet use models were adjusted for the following covariates: age, occupational status, education level, insurance status, and rurality. Covariates were selected based on univariate analysis (*P*<.10 on univariate analysis).

Figure 2. Association of markers of health literacy and health access with smartphone ownership (top) and internet use (bottom). Separate models were used for each outcome listed on the left, with smartphone ownership or internet use included as a covariate.



These differences were reflected in the sources of information that the respondents reported using to obtain health information (Figures 3 and 4). Although most of the respondents did not use physicians as their most recent source of health information, the respondents without smartphones or internet access were more likely to have used a physician for this purpose (116/349, 33.2% of nonsmartphone owners vs 344/1800, 19.11% of smartphone owners; 80/228, 35.1% of noninternet users vs 380/1921, 19.78% of internet users; both *P*<.001). Similarly, nonsmartphone owners were less likely than smartphone owners to have used the internet (96/349, 27.5% vs 1098/1800, 61%, respectively; *P*<.001) or social media (9/349, 2.6% vs 93/1800, 5.17%, respectively; *P*=.04).

Figure 3. Source used most recently for health information by smartphone ownership.

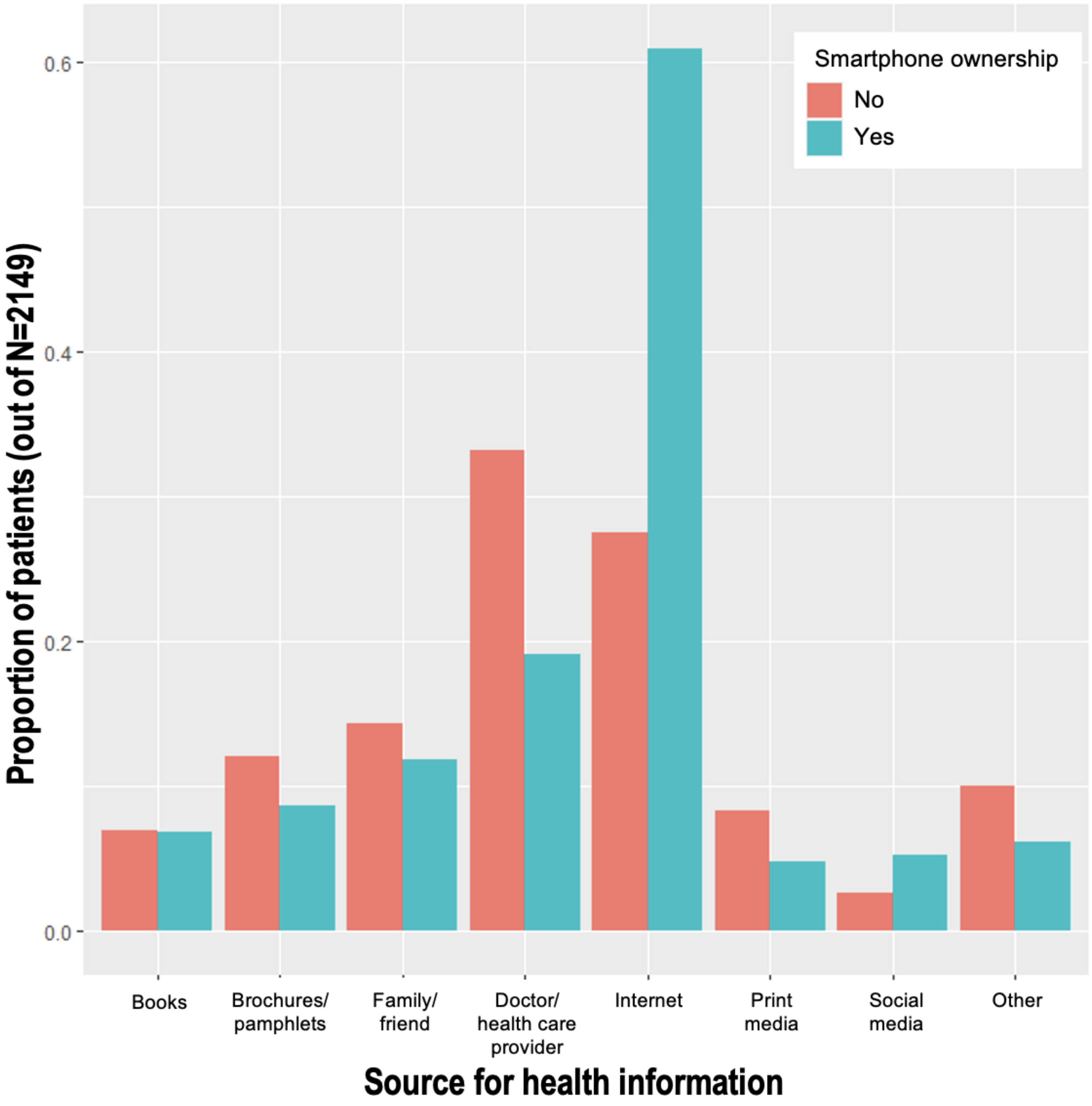
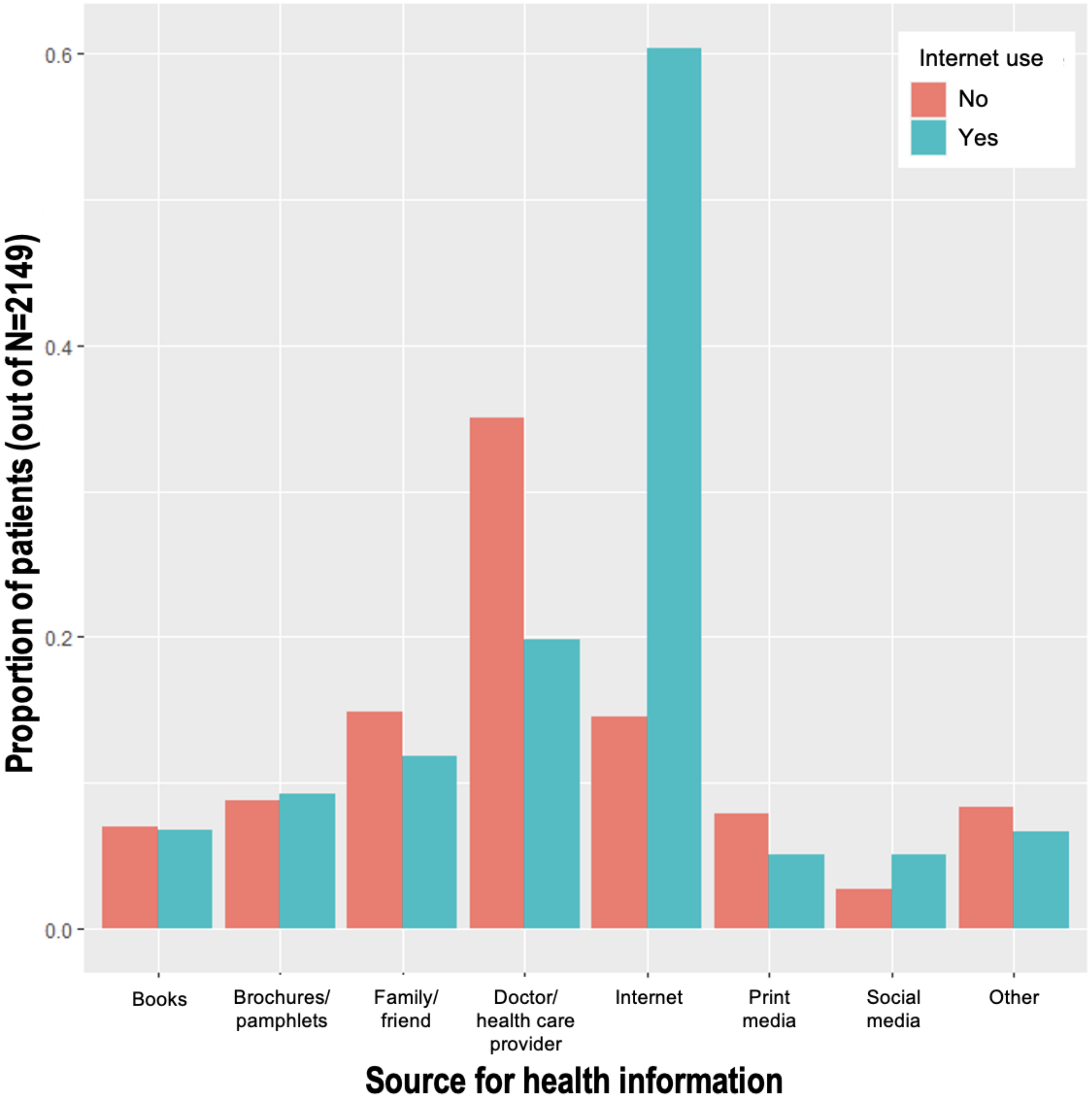


Figure 4. Source used most recently for health information by internet use.



Digital Technology Use and Health Access

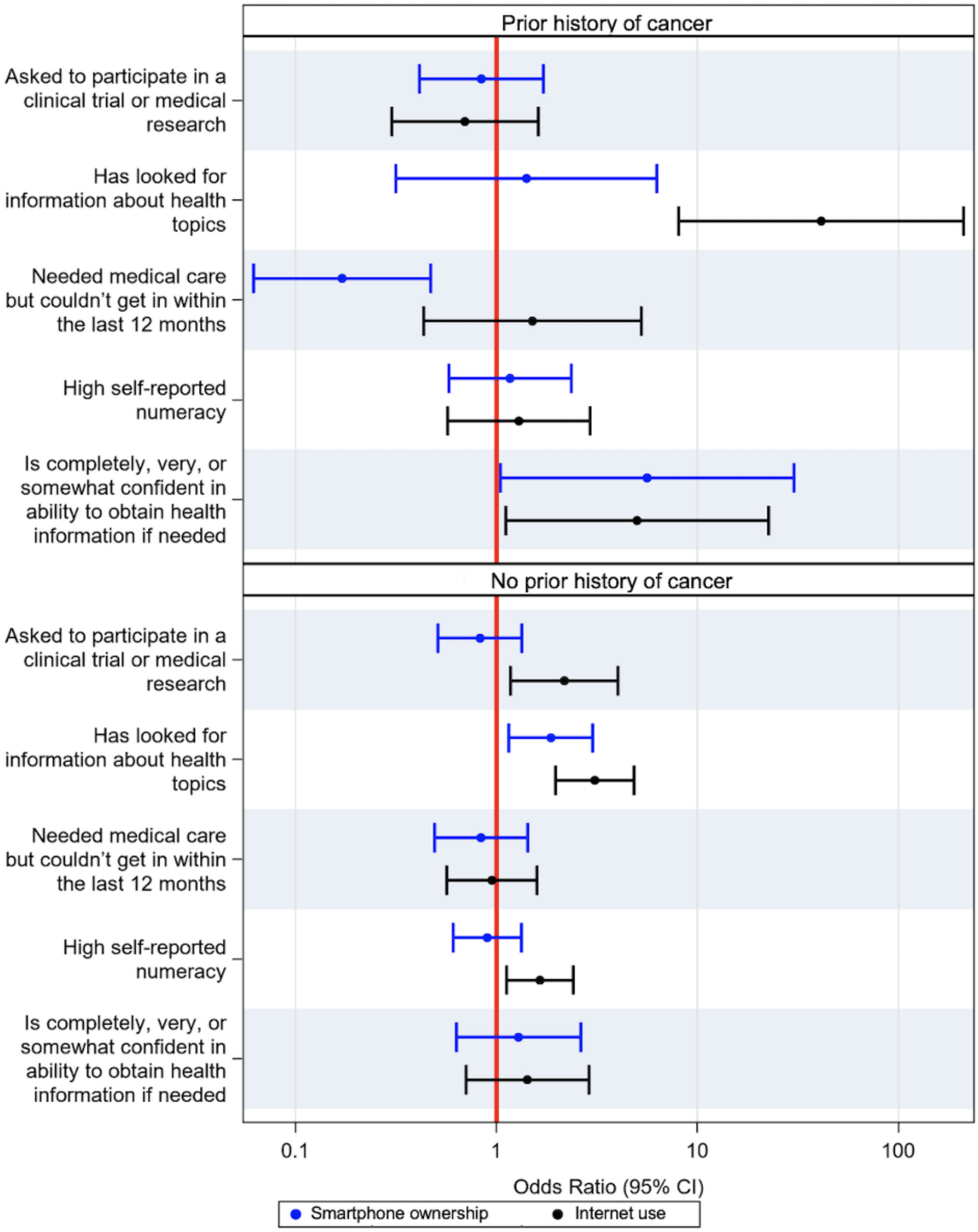
Smartphone ownership and internet use were also associated with markers of health care access (Table 2). Neither smartphone ownership nor internet use was significantly associated with having a usual place for health care or with using an outpatient office or clinic more frequently than the emergency room. However, smartphone owners were less likely than nonsmartphone owners to have been unable to obtain needed medical care within the last 12 months (234/1800, 13% vs 72/349, 20.6%, respectively; *P*<.001).

In adjusted analyses (Figure 2), controlled for age, race and ethnicity, nativity, language spoken at home, income adequacy, occupational status, education level, insurance status, and rurality, this relationship remained significant (OR 0.62, 95% CI, 0.40-0.95). On univariate analysis, internet use was associated with users having been asked to participate in medical research (73.7%, *P*=.002), and this relationship also remained significant in the adjusted analyses (OR 1.73, 95% CI 1.09-2.75) when controlling for age, occupational status, education level, insurance status, and rurality (Figure 2).

Digital Technology Use in Participants With a Personal History of Cancer

The respondents who used the internet had higher odds of a personal history of cancer (OR 1.91, 95% CI 1.21-3.01; Figure 2). In adjusted analyses, older age (OR 1.06, 95% CI 1.05-1.07), disability (OR 4.12, 95% CI 2.30-7.37), and retired status (OR 1.64, 95% CI 1.08-2.49) were associated with an increased likelihood of a personal history of cancer. After stratifying the cohort by personal history of cancer, distinct relationships among smartphone ownership, internet use, and markers of health literacy and health access were observed. For participants with a personal history of cancer, smartphone owners had higher odds of being confident in their ability to obtain health information if needed (OR 5.63, 95% CI 1.05-30.23) and lower odds of needing medical care but being unable to get it (OR 0.17, 95% CI 0.06-0.47) compared with nonsmartphone owners, whereas these associations were not maintained among participants without a personal history of cancer (Figure 5). Similarly, among the participants with a history of cancer, internet users had higher odds of being confident in their ability to obtain health information if needed (OR 5.02, 95% CI 1.12-22.55), whereas this association was not maintained among participants without a prior history of cancer. All models were adjusted for the following covariates: age, occupational status, education level, insurance status, and rurality. Covariates were selected based on univariate analysis (*P*<.10 on univariate analysis).

Figure 5. Association of smartphone ownership and internet use with markers of health literacy and health access in participants with (top) and without (bottom) a prior history of cancer. Separate models were used for each outcome listed on the left, with smartphone ownership or internet use included as a covariate.



Discussion

Principal Findings

Ensuring access to telehealth is an increasingly important priority for the medical community because the COVID-19 pandemic has shifted a large proportion of care into the digital realm, and the convenient and inexpensive nature of the medium suggests that its popularity will persist long after the pandemic eases [47]. As smartphone ownership and internet access are essential modes of digital connectivity in a world that increasingly relies upon such modalities to facilitate health care, smartphone and internet access could have significant implications for health outcomes. Prior studies describing patterns of connectivity are nearly a decade old, suggesting a gap in current evidence [48-50].

The rates of internet use, mobile phone ownership, and smartphone ownership in our cohort were high, with 89.39% (1921/2149) of the respondents reporting internet use, 96.32% (2070/2149) reporting mobile phone access, and 83.76% (1800/2149) reporting smartphone ownership. These rates are higher than prior estimates [51], highlighting the trend described in multiple studies of continued growth in digital usership [25,52,53]. Despite this increasing use, we found that a digital divide persists across socioeconomic dimensions [23,24]. Nonsmartphone and noninternet users were more likely to be older, disabled or retired, less well-educated, on public insurance, and residents of rural areas. Participants without smartphones were also more likely to report finding it difficult or very difficult to get by on their current income. These socioeconomic disparities in digital access have important implications. Both smartphone ownership and internet use were consistently associated with markers of health literacy. Importantly, our findings are the first to demonstrate an additional association between smartphone use and health care access. Finally, these data suggest that having a major medical condition, such as cancer, may serve to reverse previously described relationships between sociodemographic characteristics and technology use.

Smartphone and internet use were consistently and thematically associated with markers of health-specific and overall literacy [54]. Participants without smartphone or internet access had lower self-reported numeracy and lower confidence in their ability to obtain health information when needed. They also reported being less likely to look for information on health topics, even when controlling for potential sociodemographic confounders. These findings support an evolving body of literature that demonstrates an integral connection between digital access and health literacy. In a survey of 1077 patients at community health centers and outpatient clinics, Bailey et al [31] found that patients with adequate health literacy were more likely to own a mobile phone or smartphone and to have internet access. Similarly, a cross-sectional study of 131 low-income adults by Jensen et al [32] found that those with low health literacy skills were less likely to use the internet and related technologies. Low health literacy has been consistently associated with increased rates of hospitalization, greater use of the emergency department, increased medication nonadherence, decreased use of preventive services, and increased risk of mortality [55]. This body of data suggests that digital technology plays a key role in health outcomes.

Patient-provider communication may be an important mediator of the relationship between health literacy and health outcomes. In a survey of 823 patients presenting to an urban public hospital, Yin et al [56] found that patients with lower health literacy were more likely to rely on a doctor's knowledge to make medical decisions and less likely to rely on their own knowledge or beliefs. Similarly, the respondents in our study who did not own smartphones or use the internet were significantly more likely to use a health care provider as their most recent source of health information. Importantly, the survey study by Yin et al [56] found that patients with lower health literacy were less likely to feel like partners in shared decision-making with their providers. These findings point to the need for both provider-level and systemic changes to provide increased support for those who rely on face-to-face health communication to inform their care. In addition, these findings highlight the need for further research into provider reliance on patients' pre-existing medical knowledge or self-driven information seeking and how these behaviors may contribute to health disparities.

Our data are the first to demonstrate an association between smartphone ownership and disparities in access to care. Within our cohort, lack of smartphone ownership was significantly associated with being unable to obtain needed medical care, even when controlling for potential sociodemographic confounders, suggesting that characteristics associated with the use of digital technologies play a role in increasing health care access. Health literacy may be an important factor because prior studies have demonstrated an association between low health literacy and self-reported difficulty in accessing care [55,57]. However, health literacy is unlikely to be the only determinant because lack of internet use in our cohort was associated with lower markers of health literacy but was not significantly associated with being unable to obtain needed medical care. Taken together, these findings suggest that patients lacking smartphone access represent a unique subpopulation, distinct from those lacking internet access, and they may be the most vulnerable to the third digital divide [30] or the digital divide representing disparities in the impact of technology use on health outcomes. Indeed, even within our sample, nonsmartphone owners were distinct from noninternet users in key sociodemographic characteristics, including income adequacy. Future work should be directed toward identifying the unique barriers to care encountered by patients lacking smartphone access, including qualitative studies to assess patient experiences.

Finally, we found that a major health event such as a personal history of cancer may help overcome previously established sociodemographic patterns in digital technology use, with important implications for health outcomes. In our study, patients with a personal history of cancer were more likely to report internet use, despite being older than those without cancer, indicating that populations with historically low adoption of digital technologies, such as older patients [23], can be motivated to change their behavior when faced with a major illness. Importantly, our findings suggest that expanding digital access may also help to mitigate inequities in health care in this patient population. Among participants with a prior history of cancer, internet use and smartphone ownership were both significantly associated with confidence in being able to obtain needed health information. Smartphone use was also associated with lower odds of being unable to obtain needed medical care. In summary, the adoption of digital technologies in this patient population is associated with increased patient health activation, which in turn has been shown to have lasting benefits on health outcomes and health costs [58-60]. These findings demonstrate that engagement with the medical system in the form of a major illness can serve as a catalyst to overcome sociodemographic barriers to technology access, with the potential for improvement in long-term outcomes.

Study Limitations

Our study included limitations that should be acknowledged. First, this was a cross-sectional study, and as such no conclusions can be drawn regarding the causal relationships among the variables studied. Second, the goals of this study were mainly exploratory; therefore, no adjustments were made for multiple comparisons in the statistical analyses. Third, the study findings are derived from self-reported data, which are subject to recall bias and may not accurately reflect participant health literacy and health care use. Fourth, there was a relatively small sample size of participants with a prior history of cancer who did not engage in smartphone or internet use, limiting our goal of precisely estimating the relationship between technology use and markers of health literacy and access in this cohort. The impact of cancer history on the relationship among digital technology, health literacy, and health access may be best explored through an alternative study design, such as a case-control study. In addition, these survey data were collected in the second half of 2017, and although they capture more recent trends in digital connectivity than those available in currently published literature, patterns of digital technology use have continued to evolve since these data were collected. Finally, this study cohort represented a convenience sample in which study recruitment was purposefully targeted to allow the study of a diverse cohort that allows for in-depth analysis of previously understudied patient populations. Thus, there may have been unmeasured biases in the study population, and the extent to which our findings may be generalizable to the national or statewide population is unknown. Future work, including large-scale population-based studies and qualitative investigations, will help to elucidate the connections among digital technology use, health literacy, and health access that were explored here.

Conclusions

In conclusion, we found that access to digital technology has markedly increased across all social strata over the past decade in this diverse cohort of participants, leading to a more electronically connected society than ever before. This substantiates the claim that a broad digital infrastructure exists to support telemedicine as an increasingly important mode of patient engagement and health communication in the coming years. Nonetheless, a digital divide persists along sociodemographic and socioeconomic lines, with implications for both health literacy and access to care. Thus, dissemination of health technology must include measures to reach those with the most compromised health access. Importantly, our data support the idea that prolonged, meaningful contact with the health care system in the form of a major illness such as cancer has the potential to overcome sociodemographic trends in digital technology use and amplify the benefits these technologies confer. Future work should be directed toward the implementation of interventions to bridge the technology gap. In addition, more resources must be directed toward assisting vulnerable communities in engaging with applications of technology in the health sector because these applications hold promise in helping to mitigate health inequities and improve overall community health.

Acknowledgments

This work was supported by the Duke Cancer Institute through the National Institutes of Health grant P30CA014236 (principal investigator: Dr Michael B Kastan). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

SMT had a consultancy with Abbvie, Inc, on work related to bioequivalence. This relationship ended in January 2019 and was unrelated to the work. OMF is currently supported by the National Institutes of Health under award number 1K08CA241390 (principal investigator: OMF).

Multimedia Appendix 1
Participant survey items.
[DOCX File , 92 KB - jmir_v23i6e24947_app1.docx]

References

1. Gadzinski AJ, Gore JL, Ellimoottil C, Odisho AY, Watts KL. Implementing telemedicine in response to the COVID-19 pandemic. J Urol 2020 Jul;204(1):14-16. [doi: 10.1097/ju.0000000000001033]

2. Fankhauser GT. Delivering high-quality vascular care by telehealth during the COVID-19 pandemic. J Vasc Surg 2020 Jul;72(1):6-7 [FREE Full text] [doi: 10.1016/j.jvs.2020.04.010] [Medline: 32289437]
3. Ananthakrishnan AN, Singh S. The doctor will call you now! Telemedicine in the midst of a pandemic. Clin Gastroenterol Hepatol 2020 Jul;18(8):1688-1690 [FREE Full text] [doi: 10.1016/j.cgh.2020.04.031] [Medline: 32315781]
4. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
5. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
6. Irizarry T, Dabbs AD, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
7. Claridy MD, Hudson MM, Caplan L, Mitby PA, Leisenring W, Smith SA, et al. Patterns of internet-based health information seeking in adult survivors of childhood cancer. Pediatr Blood Cancer 2018 May 19;65(5):e26954 [FREE Full text] [doi: 10.1002/pbc.26954] [Medline: 29350454]
8. Cocco AM, Zordan R, Taylor DM, Weiland TJ, Dilley SJ, Kant J, et al. Dr Google in the ED: searching for online health information by adult emergency department patients. Med J Aust 2018 Oct 15;209(8):342-347. [doi: 10.5694/mja17.00889] [Medline: 30107763]
9. Corrales DM, Wells AE, Breitkopf CR, Pena G, Kaplan AL, King LS, et al. Internet use by gynecologic oncology patients and its relationship with anxiety. J Health Commun 2018 Feb 23;23(3):299-305. [doi: 10.1080/10810730.2018.1442529] [Medline: 29474124]
10. George G, Iwuanyanwu E, Buford A, Piha-Paul SA, Subbiah V, Fu S, et al. Cancer-related internet use and its association with patient decision making and trust in physicians among patients in an early drug development clinic: a questionnaire-based cross-sectional observational study. J Med Internet Res 2019 Mar 14;21(3):e10348 [FREE Full text] [doi: 10.2196/10348] [Medline: 30869638]
11. Loeb S, Sengupta S, Butaney M, Macaluso JN, Czarniecki SW, Robbins R, et al. Dissemination of misinformative and biased information about prostate cancer on Youtube. Eur Urol 2019 Apr;75(4):564-567. [doi: 10.1016/j.eururo.2018.10.056] [Medline: 30502104]
12. 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]
13. Shneyderman Y, Rutten LJ, Arheart KL, Byrne MM, Kornfeld J, Schwartz SJ. Health information seeking and cancer screening adherence rates. J Cancer Educ 2016 Mar 27;31(1):75-83. [doi: 10.1007/s13187-015-0791-6] [Medline: 25619195]
14. Hudson DB, Campbell-Grossman C, Hanna KM, Fial AV. Health-related mobile phone-based programs and health-care utilization by mothers of infants: an integrative review. J Technol Behav Sci 2019 Dec 20;5(2):121-128. [doi: 10.1007/s41347-019-00119-5]
15. Bendtsen M, Müssener U, Linderoth C, Thomas K. A mobile health intervention for mental health promotion among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e17208 [FREE Full text] [doi: 10.2196/17208] [Medline: 32196462]
16. Martin C, Gilmore L, Apolzan J, Myers C, Thomas D, Redman L. Smartloss: a personalized mobile health intervention for weight management and health promotion. JMIR Mhealth Uhealth 2016 Mar 16;4(1):e18 [FREE Full text] [doi: 10.2196/mhealth.5027] [Medline: 26983937]
17. Barsky J, Hunter R, McAllister C, Yeates K, Campbell N, Liu P, et al. Analysis of the implementation, user perspectives, and feedback from a mobile health intervention for individuals living with hypertension (DREAM-GLOBAL): mixed methods study. JMIR Mhealth Uhealth 2019 Dec 09;7(12):e12639 [FREE Full text] [doi: 10.2196/12639] [Medline: 31815678]
18. Report to congress: e-health and telemedicine. U.S. Department of Health and Human Services. 2016. URL: https://aspe.hhs.gov/system/files/pdf/206751/TelemedicineE-HealthReport.pdf [accessed 2021-05-15]
19. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017 Oct 19;377(16):1585-1592. [doi: 10.1056/nejmsr1503323]
20. Doolittle GC, Spaulding AO. Providing access to oncology care for rural patients via telemedicine. J Oncol Pract 2006 Sep 01;2(5):228-230. [doi: 10.1200/jop.2.5.228]
21. Uscher-Pines L, Mehrotra A. Analysis of Teladoc use seems to indicate expanded access to care for patients without prior connection to a provider. Health Aff (Millwood) 2014 Feb;33(2):258-264. [doi: 10.1377/hlthaff.2013.0989] [Medline: 24493769]
22. van Dijk JA. Digital divide research, achievements and shortcomings. Poetics 2006 Aug;34(4-5):221-235. [doi: 10.1016/j.poetic.2006.05.004]
23. Ernsting C, Dombrowski SU, Oedekoven M, Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]

24. Liu Z, Ng M, Gunasekeran DV, Li H, Ponampalam K, Ponampalam R. Mobile technology: usage and perspective of patients and caregivers presenting to a tertiary care emergency department. World J Emerg Med 2020;11(1):5-11 [FREE Full text] [doi: 10.5847/wjem.j.1920-8642.2020.01.001] [Medline: 31892997]
25. Anderson M. Mobile technology and home broadband 2019. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2020-04-27]
26. Hillyer GC, Schmitt KM, Lizardo M, Reyes A, Bazan M, Alvarez MC, et al. Electronic Communication Channel Use and Health Information Source Preferences Among Latinos in Northern Manhattan. J Community Health 2017 Apr 21;42(2):349-357 [FREE Full text] [doi: 10.1007/s10900-016-0261-z] [Medline: 27655586]
27. Smith A. Older Adults and Technology Use. Pew Research Center. 2014. URL: https://www.pewresearch.org/internet/2014/04/03/older-adults-and-technology-use/ [accessed 2021-05-15]
28. Gonzales A. The contemporary US digital divide: from initial access to technology maintenance. Inf Commun Soc 2015 Jun 02;19(2):234-248. [doi: 10.1080/1369118x.2015.1050438]
29. van Deursen AJ, van Dijk JA. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. New Media Soc 2019 Feb 07;21(2):354-375 [FREE Full text] [doi: 10.1177/1461444818797082] [Medline: 30886536]
30. Neter E, Brainin E, Baron-Epel O. The third digital divide in the health domain: is internet use for health purposes associated with health benefits? In: Hale TM, Chou WY, Cotten SR, editors. eHealth: Current Evidence, Promises, Perils, and Future Directions. Bingley, United Kingdom: Emerald Publishing Limited; 2018:1-320.
31. 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 02;18(6):3079-3087 [FREE Full text] [doi: 10.1111/hex.12294] [Medline: 25363660]
32. Jensen JD, King AJ, Davis LA, Guntzviller LM. Utilization of internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. J Aging Health 2010 Sep 21;22(6):804-826. [doi: 10.1177/0898264310366161] [Medline: 20495159]
33. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The relationship of health literacy with use of digital technology for health information: implications for public health practice. J Public Health Manag Pract 2017;23(4):380-387. [doi: 10.1097/PHH.0000000000000366] [Medline: 26672402]
34. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
35. Estacio EV, Whittle R, Protheroe J. The digital divide: examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J Health Psychol 2019 Oct 01;24(12):1668-1675. [doi: 10.1177/1359105317695429] [Medline: 28810415]
36. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
37. Neter E, Brainin E. Association between health literacy, ehealth literacy, and health outcomes among patients with long-term conditions. Eur Psychol 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
38. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]
39. Osborn R, Squires D, Doty MM, Sarnak DO, Schneider EC. In new survey of eleven countries, US adults still struggle with access to and affordability of health care. Health Aff (Millwood) 2016 Dec 01;35(12):2327-2336. [doi: 10.1377/hlthaff.2016.1088] [Medline: 27856648]
40. Mattsson S, Olsson EM, Johansson B, Carlsson M. Health-related internet use in people with cancer: results from a cross-sectional study in two outpatient clinics in Sweden. J Med Internet Res 2017 May 15;19(5):e163 [FREE Full text] [doi: 10.2196/jmir.6830] [Medline: 28506959]
41. Ginossar T. Predictors of online cancer prevention information seeking among patients and caregivers across the digital divide: a cross-sectional, correlational study. JMIR Cancer 2016 Mar 09;2(1):e2 [FREE Full text] [doi: 10.2196/cancer.5108] [Medline: 28410177]
42. Tai C, Hiatt R. The population burden of cancer: research driven by the catchment area of a cancer center. Epidemiol Rev 2017 Jan 01;39(1):108-122. [doi: 10.1093/epirev/mxx001] [Medline: 28472310]
43. Barrett N, Ingraham K, Bethea K, Hwa-Lin P, Chirinos M, Fish LJ, et al. Project PLACE: enhancing community and academic partnerships to describe and address health disparities. Adv Cancer Res 2020;146:167-188 [FREE Full text] [doi: 10.1016/bs.acr.2020.01.009] [Medline: 32241388]
44. Gage-Bouchard EA, Rawl SM. Standardizing measurement of social and behavioral dimensions of cancer prevention and control to enhance outreach and engagement in NCI-designated cancer centers. Cancer Epidemiol Biomarkers Prev 2019 Jan 22;28(3):431-434. [doi: 10.1158/1055-9965.epi-18-0794]
45. Rural-urban continuum codes. Economic Research Service, U.S. DEPARTMENT OF AGRICULTURE. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/ [accessed 2020-03-10]

46. Butler M, Beale C. Rural-urban continuum codes for metro and nonmetro counties, 1993. United States Department of Agriculture, Economic Research Service 1994:278774. [doi: 10.22004/ag.econ.278774]
47. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
48. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009 Sep 01;16(5):683-689. [doi: 10.1197/jamia.m3169]
49. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct 7;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
50. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct 3;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]
51. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. Pew Research Center. 2016. URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2021-05-15]
52. Smith A. Smartphone ownership-2013 update. Pew Research Center. 2013. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_Smartphone_adoption_2013_PDF.pdf [accessed 2021-05-15]
53. Torous J, Chan SR, Tan SY, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. JMIR Ment Health 2014 Dec 23;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
54. Kiechle ES, Bailey SC, Hedlund LA, Viera AJ, Sheridan SL. Different measures, different outcomes? A systematic review of performance-based versus self-reported measures of health literacy and numeracy. J Gen Intern Med 2015 Oct 28;30(10):1538-1546 [FREE Full text] [doi: 10.1007/s11606-015-3288-4] [Medline: 25917656]
55. 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]
56. Yin HS, Dreyer BP, Vivar KL, MacFarland S, van Schaick L, Mendelsohn AL. Perceived barriers to care and attitudes towards shared decision-making among low socioeconomic status parents: role of health literacy. Acad Pediatr 2012 Mar;12(2):117-124 [FREE Full text] [doi: 10.1016/j.acap.2012.01.001] [Medline: 22321814]
57. Levy H, Janke A. Health literacy and access to care. J Health Commun 2016 Apr 04;21 Suppl 1(sup1):43-50 [FREE Full text] [doi: 10.1080/10810730.2015.1131776] [Medline: 27043757]
58. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May 30;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
59. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015 Mar;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]
60. Hibbard JH, Greene J, Shi Y, Mittler J, Scanlon D. Taking the long view: how well do patient activation scores predict outcomes four years later? Med Care Res Rev 2015 Jun 24;72(3):324-337. [doi: 10.1177/1077558715573871] [Medline: 25716663]

Abbreviations

RUCC: Rural-Urban Continuum Code
OR: odds ratio

Edited by R Kukafka; submitted 11.10.20; peer-reviewed by E Neter, S Speedie, A Hidki, JR Bautista; comments to author 10.01.21; revised version received 08.03.21; accepted 25.04.21; published 09.06.21.

Please cite as:
Oshima SM, Tait SD, Thomas SM, Fayanju OM, Ingraham K, Barrett NJ, Hwang ES
Association of Smartphone Ownership and Internet Use With Markers of Health Literacy and Access: Cross-sectional Survey Study of Perspectives From Project PLACE (Population Level Approaches to Cancer Elimination)
J Med Internet Res 2021;23(6):e24947
URL: https://www.jmir.org/2021/6/e24947
doi:10.2196/24947
PMID:34106076

Original Paper

Digital Information Technology Use, Self-Rated Health, and Depression: Population-Based Analysis of a Survey Study on Older Migrants

Anne Kouvonen[1,2,3], MSocSci, PhD; Laura Kemppainen[1], MA, PhD; Eeva-Leena Ketonen[1], BSocSci; Teemu Kemppainen[4], MSocSci, PhD; Antero Olakivi[1], MSocSci, PhD; Sirpa Wrede[1], MSocSci, PhD

[1]Faculty of Social Sciences, University of Helsinki, Helsinki, Finland
[2]Research Institute of Psychology, SWPS University of Social Sciences and Humanities, Wroclaw, Poland
[3]Centre for Public Health, Queen's University Belfast, Belfast, United Kingdom
[4]Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland

Corresponding Author:
Anne Kouvonen, MSocSci, PhD
Faculty of Social Sciences
University of Helsinki
PO Box 54
Helsinki, 00014
Finland
Phone: 358 504487113
Email: anne.kouvonen@helsinki.fi

Abstract

Background: Previous studies have found that in general, poor health is associated with a lower likelihood of internet use in older adults, but it is not well known how different indicators of health are associated with different types of digital information technology (DIT) use. Moreover, little is known about the relationship between health and the types of DIT use in older ethnic minority and migrant populations.

Objective: The aim of this study is to examine the associations among depressive symptoms and self-rated health (SRH) with different dimensions of DIT use in older migrants.

Methods: We analyzed data from the Care, Health and Ageing of Russian-speaking Minority (CHARM) study, which is based on a nationally representative sample of community-dwelling, Russian-speaking adults aged 50 years or older residing permanently in Finland (men: 616/1082, 56.93%; age: mean 63.2 years, SD 8.4 years; response rate: 1082/3000, 36.07%). Data were collected in 2019 using a postal survey. Health was measured using depressive symptoms (measured using the Center for Epidemiologic Studies Depression Scale) and SRH. Binary logistic regression analyses were used to investigate the associations between the two health indicators and the following six outcomes: daily internet use, smartphone ownership, the use of the internet for messages and calls, social media use, the use of the internet for personal health data, and obtaining health information from the internet. A number of sociodemographic and socioeconomic factors were controlled for in the logistic regression regression analysis. Analyses were performed with weights accounting for the survey design and nonresponse.

Results: After adjusting for sociodemographic and socioeconomic factors, depressive symptoms (odds ratio [OR] 2.68, 95% CI 1.37-5.24; *P*=.004) and poor SRH (OR 7.90, 95% CI 1.88-33.11; *P*=.005) were associated with a higher likelihood of not using the internet daily. Depressive symptoms (OR 1.88, 95% CI 1.06-3.35; *P*=.03) and poor SRH (OR 5.05, 95% CI 1.58-16.19; *P*=.006) also increased the likelihood of smartphone nonuse. Depressive symptoms were additionally associated with a lower likelihood of social media use, and poor SRH was associated with a lower likelihood of using the internet for messaging and calling.

Conclusions: Poor SRH and depressive symptoms are associated with a lower likelihood of DIT use in older adults. Longitudinal studies are required to determine the directions of these relationships.

(J Med Internet Res 2021;23(6):e20988) doi:10.2196/20988

KEYWORDS

digital information technology; older adults; migrants; health; depression; mobile phone

Introduction

Background

The use of digital information technology (DIT) has become an essential and pervasive part of daily living for all generations. However, the so-called digital divide continues to exist because certain segments of the population still cannot access the latest information technologies [1]. In general, there are still large differences between age and cohort groups: although virtually all young and middle-aged adults now use the internet, a considerable proportion of older adults, especially those aged above 75 years, do not have access to the web [2]. The other determinants of DIT use include socioeconomic factors such as education, income, and poverty [3-5], as well as ethnic background [4-8] and migrant status [8]. There is a strong link between traditional social exclusion and digital exclusion [9].

In addition, health is associated with DIT use. This becomes particularly relevant in older age, when health issues can increasingly hamper daily and social activities. A large survey of a representative sample of older adults in 17 European countries showed that better self-rated health (SRH) was associated with an increased likelihood of internet use [3]. This result has been confirmed in other studies [10-19]. Similarly, lower depression levels [5,20-22], better functional capacity [14], better well-being and quality of life [23-25], healthier lifestyles [23,26], fewer chronic medical conditions [5], and more favorable cardiovascular risk factors [27] have been associated with a higher likelihood of DIT use in older adults. In contrast, some studies have found no independent association between health status and internet use [4,13].

However, previous studies included several limitations. First, internet use has typically been assessed as a binary variable, although it comprises vastly different facets and dimensions. Health may affect different aspects of DIT use in different ways. Older adults use the internet in diverse ways, and smartphone use has increased dramatically over the last 5 years; both have implications for the patterns of internet use. Indeed, studies measuring the scope and heterogeneity of DIT use by older adults have been called for [3,17,25,28,29]. Second, many studies have measured health with only one indicator; this can be problematic because studies that have used a variety of physical and mental health and functioning indicators have shown differing associations with DIT use [11].

Third, *older adults* are becoming an increasingly diverse population group. In particular, the number and proportion of older migrants are increasing rapidly in many countries. Despite of this, older migrants have typically been overlooked in studies investigating health and DIT use. Moreover, there are only limited studies on DIT use in ethnic minority older populations other than African Americans and Hispanics in the United States [6,12].

Objective

To address these limitations, the aim of this study is to examine the associations among depressive symptoms and SRH with different dimensions of DIT use in a representative sample of older Russian-speaking migrants in Finland. Theoretically, this study is placed in the framework of multidimensional and intersectional forms of digital and social exclusion. Intersecting domains such as older age, ill-health, and a migrant background create a higher risk of social exclusion [30,31]. In a rapidly digitizing society, digital exclusion is becoming an increasingly important aspect of social exclusion; in this regard, older migrants with poor health may be a particularly vulnerable group. The so-called double jeopardy hypothesis states that older migrants are in danger of facing exclusion risks because of their age and migrant background [32]. If we add poor health to the mix, older migrants with poor health can be seen as being subject to triple jeopardy. However, there are differences in terms of the type and purpose of internet use. Because older adults engage in a diverse range of internet activities, the different types of uses are likely to be influenced by different (health and other) characteristics [5].

Methods

Data

The Care, Health and Ageing of Russian-speaking Minority in Finland (CHARM) study focuses on Russian-speaking community-dwelling older adults (aged 50 years or older) who reside permanently in Finland. People born in Russia or the former Soviet Union constitute the largest migrant group in Finland, accounting for approximately 20% of the total migrant population. The study examines issues related to health and well-being, public service experiences, digital inclusion, and access to different types of care. Data were collected in 2019. A random sample of 3000 people was drawn from the Population Registry, which covers all persons registered as living in Finland. The sample was stratified by gender. Of those invited, a total of 36.07% (1082/3000) of people (men: 616/1082, 56.93%; age: mean 63.2 years, SD 8.4 years) agreed to participate in the study. The participants were asked to answer the questionnaire in Russian or Finnish. A total of 82 participants responded on the web, and the rest responded through a postal survey. The survey responses were weighted to adjust for nonresponse bias. The Finnish Tax Administration register data from 2017 were used to model response propensity. The data included information on earnings and capital income, unemployment benefits, earnings-related and national pensions, and student benefits.

Participation was voluntary, and the participants were informed of their right to withdraw at any time without any consequences. The study protocol was approved by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioural Sciences (#6/2019). The study conformed to the principles embodied in the Declaration of Helsinki.

DIT Use

The frequency of internet use was measured on a 7-point scale with response options ranging from “never” to “several times a day.” For the analysis, this variable was dichotomized; the last 2 categories (“once a day” and “several times a day”) were merged to indicate daily internet use (vs others).

A dichotomized (yes or no) follow-up question captured the different types of internet use among the users. The following items were deemed potentially the most relevant for health and were therefore included in this study: messages and calls, social media, accessing personal health data, and obtaining health information. Messages, calls, and social media use relate to social relationships and social connectedness and interaction, which are well-established determinants of health, and accessing personal health data and obtaining health information relate to a person’s health matters. Smartphone use was assessed with a dichotomous question—“Do you own a smartphone?”

Health Indicators

The CHARM questionnaire included a number of health indicators, such as depressive symptoms; SRH; doctor-diagnosed diseases; limiting long-term illness; physical functioning measured as having difficulties in walking up 3 flights of stairs and difficulties in walking approximately half a kilometer without breaks with or without a walker or walking stick; and reporting hearing, vision or memory, concentration, and learning difficulties. Of these, for this study, we selected the 2 generic indicators that most widely and comprehensively cover physical and mental health—SRH and depressive symptoms.

The 8-item Center for Epidemiologic Studies Depression Scale [33] was used to measure depressive symptoms, with a score of 9 or more points indicating depressive symptoms. The scale and cut-off have been validated for use in community-dwelling older adults [34].

SRH (“In general, would you say your health now is...?”) was assessed on a 5-point scale (1=good; 2=fairly good; 3=average; 4=fairly poor; 5=poor). The groups reporting fairly poor (n=76) or poor (n=28) SRH were merged because of the low numbers of participants in these categories. The single-item SRH measure has been shown to be a valid and reliable instrument that strongly predicts mortality [35,36].

Covariates

Sex, age, marital status (married or cohabiting vs other), educational qualification in Finland (yes or no), educational level in the country of origin, proficiency in local languages, Finnish citizenship (yes or no), the receipt of income support (yes or no), and the type of survey participation (web-based or postal survey) were included as covariates. The highest educational level in the country of origin was categorized as having no education or basic education, vocational training, or higher education. This categorization is based on the structure of the educational system in the former Soviet Union (the participants were of school age during Soviet times and completed their education mainly in that country because most had moved to Finland as adults).

Acculturation was measured by the degree to which the participant had learned the local official language and whether they had obtained Finnish citizenship. The participants assessed their proficiency in local languages on a 4-point scale. The response options “I use Finnish or Swedish language in various ways in different situations” and “I can participate on everyday conversations in Finnish or Swedish” were categorized as having good proficiency. The response options “I can cope with simple everyday situations in Finnish or Swedish” and “I do not speak either language at all” were merged to indicate poor proficiency. Income support receipt was used to measure poverty. Income support is a means-tested, last-resort financial assistance benefit in Finland [37].

Statistical Analysis

Descriptive analyses of the sample included bivariate comparisons (frequencies) of sociodemographic characteristics and health indicators (depressive symptoms and SRH) between those who used DIT and those who did not and testing for any differences between these groups (using the chi-square test). Binary logistic regression analysis was used to examine the associations between the health indicators and DIT use. We estimated odds ratios (ORs) and their 95% CIs for different types of DIT use by the health indicators by first controlling for age and sex and then further controlling for marital status, educational attainment, proficiency in local languages, Finnish citizenship, income support, and the type of participation (web-based or postal survey). Sensitivity analyses were performed on the unweighted sample. The two health indicators were not entered simultaneously into the analyses. Health indicators may partly overlap, and entering them simultaneously into the models can cause overadjustment.

The analyses were conducted using Stata, version 15.1 (StataCorp LLC).

Results

Descriptive Statistics

Table 1 shows the characteristics of the sample and the distribution of the study variables by daily internet nonuse and smartphone nonuse. Most of the participants (796/1067, 74.6%) were married or cohabiting, 39.97% (400/1082) had obtained some educational qualifications in Finland, and 50% (541/1082) had acquired a higher education in their country of origin. Half of the participants had Finnish citizenship, and 37.78% (385/1019) rated their local language (Finnish or Swedish) skills as good. Deprivation levels were very high because 41.56% (421/1013) of the participants had received income support. Of the participants, 20.08% (194/966) reported depressive symptoms. Half of the participants had average SRH, 26.24% (281/1071) saw their health as fairly good, and 14.01% (150/1071) as good, whereas 9.71% (104/1071) reported that they had fairly poor or poor SRH.

Table 1. Characteristics of the study sample and the prevalence of digital information technology nonuse.

Characteristics	Total, n[a] (%)	Internet nonuse		*P* value[b]	Smartphone nonuse		*P* value[b]
		Value, N	Value, n (%)		Value, N	Value, n (%)	
Sex				.36			.32
Female	466 (43.07)	458	61 (13.32)		404	68 (16.83)	
Male	616 (56.93)	609	76 (12.48)		561	81 (14.44)	
Age (years)				<.001			<.001
50-64	653 (60.35)	646	40 (6.19)		614	51 (8.31)	
≥65	429 (39.65)	421	97 (23.04)		351	98 (27.92)	
Married or cohabiting				.006			.18
No	271 (25.40)	265	46 (17.36)		229	44 (19.21)	
Yes	796 (74.60)	789	85 (10.77)		726	102 (14.05)	
Education in Finland				<.001			<.001
No	682 (63.03)	670	110 (16.42)		597	118 (19.77)	
Yes	400 (39.97)	397	27 (6.80)		368	31 (8.42)	
Education in the country of origin				<.001			.34
General, none, or missing	82 (7.58)	79	22 (27.85)		60	11 (18.33)	
Vocational	459 (42.42)	452	73 (16.15)		405	71 (17.53)	
Higher	541 (50)	536	42 (7.84)		500	67 (13.40)	
Proficiency in local languages				<.001			<.001
Poor	634 (62.22)	623	102 (16.37)		562	110 (19.57)	
Good	385 (37.78)	382	23 (6.02)		353	30 (8.50)	
Finnish citizenship				<.001			.001
No	546 (50.98)	538	94 (17.47)		478	96 (20.08)	
Yes	525 (49.02)	519	39 (7.51)		479	49 (10.23)	
Income support				<.001			<.001
Yes	421 (41.56)	415	71 (17.11)		362	77 (21.27)	
No	592 (58.44)	583	43 (7.38)		554	61 (11.01)	
Type of participation				<.001			.05
Postal survey	1000 (92.42)	985	136 (13.81)		884	144 (16.29)	
Web-based survey	82 (7.58)	82	1 (1.22)		81	5 (6.17)	
Depression				.003			.02
Yes	194 (20.08)	189	36 (19.05)		169	39 (23.08)	
No	772 (79.92)	764	67 (8.77)		717	93 (12.97)	
Self-rated health				<.001			<.001
Fairly poor or poor	104 (9.71)	102	24 (23.52)		87	31 (35.63)	
Average	536 (50.05)	529	80 (15.12)		463	76 (16.41)	
Fairly good	281 (26.24)	277	23 (8.30)		265	32 (12.08)	
Good	150 (14.01)	150	7 (4.67)		141	7 (4.96)	

[a] N varies by variable depending on the number of missing values.

[b] Differences in the prevalence of digital information technology nonuse between groups (from two-tailed chi-square test). *P* values are weighted.

Of the 1082 participants, only 137 (12.66%) did not use the internet daily. Daily internet nonuse and smartphone nonuse did not differ between men and women. Both internet nonuse and smartphone nonuse were much more prevalent among individuals aged 65 years or older (internet nonuse: 97/421, 23%; smartphone nonuse: 98/351, 27.9%) than in the younger age group (internet nonuse: 40/646, 6.2%; smartphone nonuse: 51/614, 8.3%). However, it is notable that for the smartphone use question, the number of missing values was very high, particularly in the older age group (78/429, 18.2%). This may be due to the formulation of the question, which required the participant to circle "Yes" if they owned a smartphone and "No" if they did not. The same question included 4 other items: a tablet, other computer, safety bracelet or safety phone, and web-based banking ID. Some of the participants had only ticked "Yes" to mark the items that they possessed but failed to tick "No" for those that they did not possess. We conducted a sensitivity analysis by recoding these cases as "No" for smartphone use, which increased the prevalence of nonuse to 13.8% (90/653) in working-age older adults and to 41% (176/429) in those aged 65 years or older. It is likely that the actual prevalence of smartphone nonuse is therefore somewhere between 7.8% (51/653) and 13.8% (90/653) in those aged 50-64 years and between 22.8% (98/429) and 41% (176/429) in those aged 65 years or older.

Participants who were married or cohabiting were more likely to use the internet daily, but there was no difference between these groups in terms of smartphone use. Similarly, those who had obtained qualifications in Finland, those who had obtained higher education in their country of origin, those with good local language skills, and those with Finnish citizenship were more likely to use the internet and own a smartphone. In addition, poverty, measured as income support receipt, increased the prevalence of DIT nonuse.

Both daily internet nonuse and smartphone nonuse were more common among the participants who reported depressive symptoms: out of the 189 participants, 36 (20.1%) with depressive symptoms did not use the internet daily compared with 8.8% (67/764) of the participants with no depressive symptoms. The corresponding figures for smartphone ownership were 23.1% (39/169) and 12.9% (93/717), respectively. A similar picture emerged for SRH: compared with those in good health, those with less than good health were more often nonusers of DIT, and the poorer the health, the more common the nonuse.

Of the different types of internet use, the most common was using the internet for calls and messages (n=855). In addition, using the internet to search for health information (n=650) and social media use (n=631) were common. Compared with these, using the internet to view personal health data was much less commonly reported (n=353).

Health and DIT Use

Table 2 shows the results of the logistic regression models examining the associations between health indicators and daily internet nonuse, and Table 3 presents the associations between health indicators and smartphone nonuse. As shown in Tables 2 and 3, after full adjustments, depressive symptoms were associated with both daily internet nonuse (OR 2.68, 95% CI 1.37-5.24; *P*=.004; Table 2) and smartphone nonuse (OR 1.88, 95% CI 1.06-3.35; *P*=.03; Table 3). Similarly, poor SRH was associated with daily internet nonuse (OR 7.90, 95% CI 1.88-33.11; *P*=.005; Table 2) and smartphone nonuse (OR 5.05, 95% CI 1.58-16.19; *P*=.006; Table 3). Compared with those with good SRH, those with fairly good or average SRH also reported more daily internet nonuse, but there were no associations between these categories and smartphone nonuse.

Table 2. Associations between health indicators and daily internet nonuse.

Health indicator	Model 1[a]		Model 2[b]	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value
Depression				
No	1.00 (reference)	N/A[d]	1.00 (reference)	N/A
Yes	2.37 (1.28-4.38)	.006	2.68 (1.37-5.24)	.004
Self-rated health				
Good	1.00 (reference)	N/A	1.00 (reference)	N/A
Fairly good	6.05 (1.62-22.55)	.007	5.44 (1.50-19.81)	.01
Average	6.62 (1.86-23.60)	.004	4.77 (1.32-17.26)	.02
Fairly poor or poor	9.66 (2.40-38.81)	.001	7.90 (1.88-33.11)	.005

[a] Adjusted for sex and age.

[b] Additionally adjusted for marital status, education in Finland, education in country of origin, proficiency in local languages, citizenship, income support, and type of participation.

[c] OR: odds ratio.

[d] N/A: not applicable.

Table 3. Associations between health indicators and smartphone nonuse.

Health indicator	Model 1[a]		Model 2[b]	
	OR[c] (95% CI)	*P* value	OR (95% CI)	*P* value
Depression				
No	1.00 (reference	N/A[d]	1.00 (reference)	N/A
Yes	1.79 (1.02-3.15)	.04	1.88 (1.06-3.35)	.03
Self-rated health				
Good	1.00 (reference)	N/A	1.00 (reference)	N/A
Fairly good	1.38 (0.49-3.9	.55	1.28 (0.44-3.70)	.65
Average	2.12 (0.78-5.71)	.14	1.74 (0.62-4.85)	.29
Fairly poor or poor	6.06 (2.00-18.41)	.002	5.05 (1.58-16.19)	.006

[a] Adjusted for sex and age.

[b] Additionally adjusted for marital status, education in Finland, education in country of origin, proficiency in local languages, citizenship, income support, and type of participation.

[c] OR: odds ratio.

[d] N/A: not applicable.

Table 4 and Table 5 display the associations between health indicators and types of internet use among those participants who reported using the internet. Depressive symptoms were associated with a higher likelihood of social media nonuse (OR 1.93, 95% CI 1.24-2.98; *P*=.003), and poor SRH was associated with a lower likelihood of using the internet for calls and messages (OR 5.27, 95% CI 1.50-18.55; *P*=.01). In terms of the other types of internet use, there were no differences by depressive symptoms or SRH.

Table 4. Associations between health indicators and types of internet nonuse: messages and calls and social media among all internet users.

Health indicator	Messages and calls nonuse				Social media nonuse			
	Cases		OR[a,b] (95% CI)	*P* value	Cases		OR[b] (95% CI)	*P* value
	Value, N	Value, n (%)			Value, N	Value, n (%)		
Depression								
No	652	68 (10.43)	1.00 (reference)	N/A[c]	670	224 (33.43)	1.00 (reference)	N/A
Yes	169	32 (18.94)	1.43 (0.78-2.61)	.25	167	75 (44.91)	1.93 (1.24-2.98)	.003
Self-rated health								
Good	124	6 (4.84)	1.00 (reference)	N/A	124	31 (25)	1.00 (reference)	N/A
Fairly good	239	24 (10.04)	2.28 (0.72-7.25)	.16	244	92 (37.70)	1.30 (0.71-2.37)	.40
Average	424	54 (12.74)	1.64 (0.56-4.85)	.37	433	157 (36.26)	1.24 (0.69-2.22)	.47
Fairly poor or poor	80	19 (23.75)	5.27 (1.50-18.55)	.01	79	34 (43.04)	1.67 (0.81-3.43)	.17

[a] OR: odds ratio.

[b] Adjusted for sex, age, marital status, education in Finland, education in the home country, proficiency in local languages, citizenship, income support, and type of participation.

[c] N/A: not applicable.

Table 5. Associations between health indicators and types of internet nonuse: personal health data and health information among all internet users.

Health indicator	Personal health data nonuse				Health information nonuse			
	Cases		OR[a,b] (95% CI)	*P* value	Cases		OR[b] (95% CI)	*P* value
	Value, N	Value, n (%)			Value, N	Value, n (%)		
Depression								
No	646	410 (63.47)	1.00 (reference)	N/A[c]	663	228 (34.39)	1.00 (reference)	N/A
Yes	162	94 (58.02)	0.83 (0.54-1.27)	.39	166	53 (31.93)	1.09 (0.69-1.72)	.72
Self-rated health								
Good	120	74 (61.67)	1.00 (reference)	N/A	120	41 (34.17)	1.00 (reference)	N/A
Fairly good	234	149 (63.68)	0.96 (0.55-1.65)	.87	242	74 (30.58)	0.64 (0.36-1.14)	.13
Average	415	257 (61.93)	0.77 (0.46-1.29)	.32	427	146 (34.19)	0.68 (0.39-1.18)	.17
Fairly poor or poor	75	47 (62.67)	0.83 (0.40-1.71)	.61	78	26 (33.33)	0.71 (0.33-1.53)	.38

[a] OR: odds ratio.

[b] Adjusted for sex, age, marital status, education in Finland, education in the home country, proficiency in local languages, citizenship, income support, and type of participation.

[c] N/A: not applicable.

The results for all covariates in the full models are shown in Multimedia Appendices 1 and 2. The results from the sensitivity analyses using the unweighted sample (Multimedia Appendix 3) were mainly in the same direction as the results obtained with the weighted sample. However, in the unweighted sample, depressive symptoms were associated with a lower likelihood of using the internet for messages and calls, and poor SRH was associated with a lower likelihood of social media use. In addition, average SRH was associated with a higher likelihood of using the internet for personal health data.

Discussion

Principal Findings

The aim of this study is to examine the associations between mental and physical health and different facets of DIT use among older Russian-origin migrants in Finland and to determine whether these associations are independent of sociodemographic and socioeconomic factors. Our study confirms that DIT use is now strongly embedded in the everyday lives of older adults: the prevalence of internet use was very high in our sample, with 93.8% (606/646) of adults aged below 65 years and 76.9% (324/421) of those aged 65 years or older using the internet daily. In 2019, Finland scored the highest among all European Union countries on the Digital Economy and Society Index, which is a composite index that summarizes relevant indicators of digital performance and includes the components of connectivity, human capital, use of internet services, integration of digital technology, and digital public services [38].

In our study, depressive symptoms were associated with a higher likelihood of daily internet and smartphone nonuse. These results are in line with earlier studies that have shown an association between internet use and depression in older adults [5,20-22]. A large prospective study among US older adults showed that one pathway explaining this association is that internet use influences depression by decreasing loneliness and social isolation [20]. It is well established that loneliness is associated with depression [39], and DIT use has been shown to protect older adults in poor health from social exclusion [40]. A previous study showed that the relationship between internet use and a higher quality of life was mediated by reduced loneliness [24].

In this study, depressive symptoms were additionally associated with a higher likelihood of social media nonuse. A previous study found that social media can become an increasingly important source of feelings of connectedness in older age when retirement and declining health decrease other forms of social engagement [16]. DIT use can considerably increase older migrants' ability to maintain and expand their dispersed support networks [41]. However, mental health problems such as depression may also result in decreased initiative to engage in new activities [42] such as social media use.

Similar to previous studies [3,10-19], we found an association between SRH and internet use. In addition, poor SRH was associated with smartphone nonuse. Physical changes such as reduced visual acuity, declining motor skills, and age-related changes in cognitive abilities are likely to affect the ability to use DIT [43], and these changes may be particularly critical in terms of hindering smartphone use.

DIT provides many specific health-related resources for those with health problems. It enables them to search for health information, communicate with their health care providers, access their laboratory results, renew their prescriptions, and seek peer support. Therefore, it could be expected that those with poorer health would be more likely to use the internet to seek health information and access personal health data. Indeed, an earlier study showed that having more chronic conditions increased the odds of internet use for health-related tasks [5]. However, in our study, depression and poor SRH did not have an effect on the search for health information among those who used the internet. This finding is consistent with that of a previous Australian study [15]. Moreover, we did not find an association between health status and the use of the internet to access personal health data.

Overall, our findings suggest that in older migrants, the digital divide still exists not only between adults aged 50-64 years and those aged 65 years or older but also between those with poor mental and physical health and those with good health. It is notable that this divide was evident even in our sample, which consisted of cognitively relatively well-functioning older adults who were able to respond to a lengthy questionnaire (20 pages). In fact, it has been noted that as more and more people access the web, the social gap between the majority (the digitally included) and the minority (those who are still digitally excluded) widens [14]. It seems that those who are digitally excluded therefore become even more socially excluded than before. Indeed, a study conducted in the United Kingdom and Sweden showed that now, when access to the internet is widespread and the proportion of those without internet access has become smaller, nonuse has become more concentrated in the most vulnerable groups. In other words, nonusers in both countries increasingly consist of the most vulnerable segments of the population, that is, older participants, those with the lowest amount of education, those with the poorest health, and those who are most socially isolated [9]. Helsper and Reidolf [9] therefore argue that we now see the emergence of a digital underclass. Older migrants may be particularly vulnerable in this regard because older age, poor health, low socioeconomic status, and a migrant background as intersecting domains increase the risk of social exclusion [30,31].

In our study, we were not able to investigate the mechanisms underlying the associations between health and DIT use. However, based on other studies, it can be assumed that functional limitations that can accompany health problems may pose barriers to DIT use [5,14]. It seems that decreased social isolation and loneliness may be one pathway to alleviating depression [20].

Strengths and Limitations

Our study included several strengths. First, we considered different dimensions of health. Second, we examined DIT use in a diverse way, including daily internet use, smartphone use, and different types of internet use as outcomes. Third, we were able to adjust for a number of relevant sociodemographic and socioeconomic factors that can affect both health status and DIT use.

However, this study also included some limitations. First, our data were cross-sectional; therefore, the associations found cannot be interpreted as causal relationships. It is likely that the relationship between health and DIT use is bidirectional, additive, and synergistic [14]: good health can increase DIT use, and DIT can improve health and well-being. In a large prospective study using a variety of methods, including matching, internet use was shown to reduce depression, and the effect was the largest for older people who lived alone [20].

Second, all our DIT use and health measures were self-reported, which can cause recall bias and misclassification. Third, our results cannot be generalized to the general population or to other migrant populations. In an earlier study on working-age migrants in Finland, Russian-origin migrants were more transnationally oriented than the other 2 migrant groups—Somalis and Kurds—which is likely to manifest as a higher level of DIT use [44]. Notwithstanding these limitations, this study is an important contribution to the growing literature on health and DIT use in older adults. This study provides results that will illuminate the design of future research to incorporate technology use for a better understanding of the mental and physical health situation of vulnerable populations.

Conclusions

In Finland, as in other Western countries, DIT use is nearing saturation among older age groups, but digital divides still exist. In this study on older migrants, we found that poor SRH and depressive symptoms were associated with a lower likelihood of DIT use. Older adults with poorer health and limited mobility are often the most socially isolated, and they would particularly benefit from the diverse use of DIT [5].

Medeiros et al [14] have proposed that because DIT use such as exchange of web-based messages demands adequate social and executive functioning, it could be used as a marker of older adults' functional capacity and depression. Therefore, the monitoring of functional capacity and depression could be done routinely by health professionals during health checkups by simply asking about the patients' DIT use. However, more importantly, there is an opportunity for the use of big data because this monitoring could be done with artificial intelligence using pattern recognition, and the results could be used in health policy planning [14]. The present results also open up possible health interventions that are now particularly relevant considering the COVID-19 pandemic circumstances.

Longitudinal studies are needed to elucidate the mechanisms underlying the associations between health and DIT use. These types of studies can better inform digital inclusion policies for migrants and older adults. Evidence-based digital inclusion strategies and policies are urgently needed to prevent increasing social exclusion of the digitally excluded populations who are also in other ways the most vulnerable, given that public services are rapidly becoming mainly or solely digitally accessible.

Acknowledgments

This study was funded by the Strategic Research Council at the Academy of Finland (grants 327145 and 327148 for the DigiIn Project), the Academy of Finland (grants 312310 and 336669 for the Centre of Excellence in Research on Ageing and Care Research Group 3 Migration, Care and Ageing), Faculty of Social Sciences at the University of Helsinki (contribution to the Centre of Excellence in Research on Ageing and Care funding), Kone Foundation (Project Crossing Borders for Health and Well-Being), and the University of Helsinki 3-year research grants. The sponsors did not have any involvement in the review or approval of the manuscript for publication.

Authors' Contributions

AK and LK designed and conceptualized this study. AK directed the implementation of this study, led all aspects of the work, and drafted the manuscript. LK, TK, and ELK performed data analysis. AK, LK, AO, and SW contributed to data acquisition. AK, LK, and TK designed this study's analytic strategy. LK, ELK, TK, AO, and SW contributed to interpreting the findings, reviewing the manuscript, and revising it critically for important intellectual content. All authors approved the submission of the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full models with depression as the main predictor.
[DOCX File , 16 KB - jmir_v23i6e20988_app1.docx]

Multimedia Appendix 2
Full models with self-rated health as the main predictor.
[DOCX File , 16 KB - jmir_v23i6e20988_app2.docx]

Multimedia Appendix 3
Coefficients for depression and self-rated health and fully adjusted models without weights.
[DOCX File , 15 KB - jmir_v23i6e20988_app3.docx]

References

1. Compaine B. Falling through the net: defining the digital divide. In: The Digital Divide: Facing a Crisis or Creating a Myth?. Cambridge, Massachusetts: The MIT Press; 2001:17-46.
2. Internet Use in the UK Annual Estimates by Age, Sex, Disability and Geographical Location. Office for National Statistics. 2019. URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2020 [accessed 2021-05-30]
3. König R, Seifert A, Doh M. Internet use among older Europeans: an analysis based on SHARE data. Univ Access Inf Soc 2018 Jan 19;17(3):621-633. [doi: 10.1007/s10209-018-0609-5]
4. Arcury TA, Sandberg JC, Melius KP, Quandt SA, Leng X, Latulipe C, et al. Older adult internet use and ehealth literacy. J Appl Gerontol 2020 Feb;39(2):141-150 [FREE Full text] [doi: 10.1177/0733464818807468] [Medline: 30353776]
5. Choi NG, Dinitto DM. Internet use among older adults: association with health needs, psychological capital, and social capital. J Med Internet Res 2013 May 16;15(5):e97 [FREE Full text] [doi: 10.2196/jmir.2333] [Medline: 23681083]
6. Werner JM, Carlson M, Jordan-Marsh M, Clark F. Predictors of computer use in community-dwelling, ethnically diverse older adults. Hum Factors 2011 Oct;53(5):431-447 [FREE Full text] [doi: 10.1177/0018720811420840] [Medline: 22046718]
7. Gordon NP, Hornbrook MC. Older adults' readiness to engage with eHealth patient education and self-care resources: a cross-sectional survey. BMC Health Serv Res 2018 Mar 27;18(1):220 [FREE Full text] [doi: 10.1186/s12913-018-2986-0] [Medline: 29587721]
8. Nguyen A, Mosadeghi S, Almario CV. Persistent digital divide in access to and use of the internet as a resource for health information: Results from a California population-based study. Int J Med Inform 2017 Jul;103:49-54. [doi: 10.1016/j.ijmedinf.2017.04.008] [Medline: 28551001]
9. Helsper EJ, Reisdorf BC. The emergence of a 'digital underclass' in Great Britain and Sweden: changing reasons for digital exclusion. New Media Soc 2016 Mar 3;19(8):1253-1270. [doi: 10.1177/1461444816634676]
10. Hong YA, Cho J. Has the digital health divide widened? Trends of health-related internet use among older adults from 2003 to 2011. J Gerontol B Psychol Sci Soc Sci 2017 Sep 1;72(5):856-863. [doi: 10.1093/geronb/gbw100] [Medline: 27558403]
11. Levine DM, Lipsitz SR, Linder JA. Changes in everyday and digital health technology use among seniors in declining health. J Gerontol A Biol Sci Med Sci 2018 Mar 14;73(4):552-559. [doi: 10.1093/gerona/glx116] [Medline: 28605446]
12. Yoon H, Jang Y, Xie B. Computer use and computer anxiety in older Korean Americans. J Appl Gerontol 2016 Sep 9;35(9):1000-1010. [doi: 10.1177/0733464815570665] [Medline: 25698717]
13. Gracia E, Herrero J. Internet use and self-rated health among older people: a national survey. J Med Internet Res 2009 Dec 2;11(4):e49 [FREE Full text] [doi: 10.2196/jmir.1311] [Medline: 19955041]
14. Medeiros FDL, Xavier AJ, Schneider IJ, Ramos LR, Sigulem D, d'Orsi E. Digital inclusion and functional capacity of older adults living in Florianópolis, Santa Catarina, Brazil (EpiFloripa 2009-2010). Rev Bras Epidemiol 2012 Mar;15(1):106-122 [FREE Full text] [doi: 10.1590/s1415-790x2012000100010] [Medline: 22450497]

15. Burns P, Jones SC, Caputi P, Iverson D. Are older Australians with chronic diseases online? Health Promot J Austr 2018 Apr;29(1):72-78. [doi: 10.1002/hpja.5] [Medline: 29700931]
16. Yu RP, McCammon RJ, Ellison NB, Langa KM. The relationships that matter: social network site use and social wellbeing among older adults in the United States of America. Ageing Soc 2015 Jul 2;36(9):1826-1852. [doi: 10.1017/s0144686x15000677]
17. Falk Erhag H, Ahlner F, Rydberg Sterner T, Skoog I, Bergström A. Internet use and self-rated health among Swedish 70-year-olds: a cross-sectional study. BMC Geriatr 2019 Dec 23;19(1):365 [FREE Full text] [doi: 10.1186/s12877-019-1392-8] [Medline: 31870313]
18. Hong YA, Zhou Z, Fang Y, Shi L. The digital divide and health disparities in China: evidence from a national survey and policy implications. J Med Internet Res 2017 Sep 11;19(9):e317 [FREE Full text] [doi: 10.2196/jmir.7786] [Medline: 28893724]
19. Matthews K, Nazroo J, Marshall A. Digital inclusion in later life: cohort changes in internet use over a ten-year period in England. Age Soc 2018 Aug 2;39(9):1914-1932. [doi: 10.1017/s0144686x18000326]
20. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among retired older adults in the United States: a longitudinal analysis. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):763-771. [doi: 10.1093/geronb/gbu018] [Medline: 24671896]
21. Hamer M, Stamatakis E. Prospective study of sedentary behavior, risk of depression, and cognitive impairment. Med Sci Sports Exerc 2014 Apr;46(4):718-723 [FREE Full text] [doi: 10.1249/MSS.0000000000000156] [Medline: 24121248]
22. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among older adults. Comput Hum Behav 2012 Mar;28(2):496-499. [doi: 10.1016/j.chb.2011.10.021]
23. Duplaga M, Szulc K. The association of internet use with wellbeing, mental health and health behaviours of persons with disabilities. Int J Environ Res Public Health 2019 Sep 4;16(18):3252 [FREE Full text] [doi: 10.3390/ijerph16183252] [Medline: 31487949]
24. Khalaila R, Vitman-Schorr A. Internet use, social networks, loneliness, and quality of life among adults aged 50 and older: mediating and moderating effects. Qual Life Res 2018 Feb;27(2):479-489. [doi: 10.1007/s11136-017-1749-4] [Medline: 29210015]
25. Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. Cyberpsychol Behav Soc Netw 2015 May;18(5):268-272. [doi: 10.1089/cyber.2014.0549] [Medline: 25919967]
26. Xavier AJ, d'Orsi E, Wardle J, Demakakos P, Smith SG, von Wagner C. Internet use and cancer-preventive behaviors in older adults: findings from a longitudinal cohort study. Cancer Epidemiol Biomarkers Prev 2013 Nov;22(11):2066-2074 [FREE Full text] [doi: 10.1158/1055-9965.EPI-13-0542] [Medline: 24148971]
27. Fox CS, Hwang S, Nieto K, Valentino M, Mutalik K, Massaro JM, et al. Digital connectedness in the Framingham heart study. JAHA 2016 Apr 3;5(4):-. [doi: 10.1161/jaha.116.003193] [Medline: 27076568]
28. Olsson T, Samuelsson U, Viscovi D. At risk of exclusion? Degrees of ICT access and literacy among senior citizens. Inf Commun Soc 2019;22(1):55-72. [doi: 10.1080/1369118x.2017.1355007]
29. van Boekel LC, Peek ST, Luijkx KG. Diversity in older adults' use of the internet: identifying subgroups through latent class analysis. J Med Internet Res 2017 May 24;19(5):e180 [FREE Full text] [doi: 10.2196/jmir.6853] [Medline: 28539302]
30. Scharf T, Keating N. Conceptualising social inclusion. In: From Exclusion to Inclusion in Old Age: A Global Challenge. New York, USA: The Policy Press; 2012:1-16.
31. Torres S. International migration: patterns and implications for exlusion in old age. In: From Exclusion to Inclusion in Old Age: A Global Challenge. New York, USA: The Policy Press; 2012.
32. Blakemore K, Boneham M. Age, Race & Ethnicity: A Comparative Approach. Buckingham, UK: Open University Press; 1994.
33. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Measure 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
34. Briggs R, Carey D, O'Halloran A, Kenny R, Kennelly S. Validation of the 8-item centre for epidemiological studies depression scale in a cohort of community-dwelling older people: data from the Irish longitudinal study on ageing (TILDA). Eur Geriatr Med 2018 Jan 3;9(1):121-126. [doi: 10.1007/s41999-017-0016-0]
35. DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality prediction with a single general self-rated health question. A meta-analysis. J Gen Intern Med 2006 Mar;21(3):267-275 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00291.x] [Medline: 16336622]
36. Bombak AE. Self-rated health and public health: a critical perspective. Front Public Health 2013;1:15 [FREE Full text] [doi: 10.3389/fpubh.2013.00015] [Medline: 24350184]
37. Social Assistance. Kela Social Insurance Institution. URL: https://www.kela.fi/web/en/social-assistance [accessed 2021-05-25]
38. The Digital Economy and Society Index (DESI). European Commission. 2019. URL: https://ec.europa.eu/digital-single-market/en/desi [accessed 2021-05-25]
39. Cacioppo JT, Hughes ME, Waite LJ, Hawkley LC, Thisted RA. Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. Psychol Aging 2006 Mar;21(1):140-151. [doi: 10.1037/0882-7974.21.1.140] [Medline: 16594799]

40. Sacker A, Ross A, MacLeod CA, Netuveli G, Windle G. Health and social exclusion in older age: evidence from Understanding Society, the UK household longitudinal study. J Epidemiol Community Health 2017 Jul 22;71(7):681-690 [FREE Full text] [doi: 10.1136/jech-2016-208037] [Medline: 28228466]
41. Millard A, Baldassar L, Wilding R. The significance of digital citizenship in the well-being of older migrants. Public Health 2018 May;158:144-148. [doi: 10.1016/j.puhe.2018.03.005] [Medline: 29625734]
42. Slegers K, van Boxtel MP, Jolles J. Computer use in older adults: determinants and the relationship with cognitive change over a 6year episode. Comput Hum Behav 2012 Jan;28(1):1-10. [doi: 10.1016/j.chb.2011.08.003]
43. Czaja SJ, Lee CC. The impact of aging on access to technology. Univ Access Inf Soc 2007 Dec 8;5(4):341-349. [doi: 10.1007/s10209-006-0060-x]
44. Kemppainen T, Kemppainen L, Kuusio H, Rask S, Saukkonen P. Multifocal integration and marginalisation: a theoretical model and an empirical study on three immigrant groups. Sociology 2020 Mar 30;54(4):782-805. [doi: 10.1177/0038038520904715]

Abbreviations

CHARM: Care, Health and Ageing of Russian-speaking Minority in Finland
DIT: digital information technology
OR: odds ratio
SRH: self-rated health

Edited by R Kukafka, G Eysenbach; submitted 03.06.20; peer-reviewed by R König, A Errazuriz; comments to author 11.07.20; revised version received 28.08.20; accepted 19.05.21; published 14.06.21.

Please cite as:
Kouvonen A, Kemppainen L, Ketonen EL, Kemppainen T, Olakivi A, Wrede S
Digital Information Technology Use, Self-Rated Health, and Depression: Population-Based Analysis of a Survey Study on Older Migrants
J Med Internet Res 2021;23(6):e20988
URL: https://www.jmir.org/2021/6/e20988
doi:10.2196/20988
PMID:34125069

Short Paper

Neighborhood Broadband and Use of Telehealth Among Older Adults: Cross-sectional Study of National Survey Data Linked With Census Data

Safiyyah M Okoye[1], MSN, PhD; John F Mulcahy[1], MSPH; Chanee D Fabius[1], MA, PhD; Julia G Burgdorf[1], PhD; Jennifer L Wolff[1], PhD

Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Safiyyah M Okoye, MSN, PhD
Department of Health Policy and Management
Bloomberg School of Public Health
Johns Hopkins University
Hampton House, Room 697
624 N Broadway
Baltimore, MD, 21205
United States
Phone: 1 667 401 7149
Email: sokoye1@jhmi.edu

Abstract

Background: The COVID-19 pandemic has amplified the role of telehealth in health care delivery. Regional variation in internet access and telehealth use are well-documented, but the effect of neighborhood factors, including the pervasiveness of broadband internet, on older adults' telehealth usage in the context of internet access is not known.

Objective: This study aimed to investigate how individual and neighborhood characteristics, including the pervasiveness of neighborhood broadband internet subscription, are associated with engagement in telehealth among older adults with internet access.

Methods: In this cross-sectional study, we included 5117 community-living older adults aged ≥65 years, who participated in the 2017 National Health and Aging Trends Study with census tract–level data for participants' places of residence from the American Community Survey.

Results: Of an estimated 35.3 million community-living older adults, 21.1 million (59.7%) were internet users, and of this group, more than one-third (35.8%) engaged in telehealth. In a multivariable regression model adjusted for individual- and neighborhood-level factors, age, education, income, and the pervasiveness of neighborhood broadband internet subscription were associated with engagement in telehealth, while race, health, county metropolitan status, and neighborhood social deprivation were not. Among internet users, living in a neighborhood at the lowest (versus highest) tertile of broadband internet subscription was associated with being 40% less likely to engage in telehealth (adjusted odds ratio 0.61, 95% CI 0.42-0.87), all else equal.

Conclusions: Neighborhood broadband internet stands out as a mutable characteristic that is consequential to engagement in telehealth.

(J Med Internet Res 2021;23(6):e26242) doi:10.2196/26242

KEYWORDS

aging; broadband internet; neighborhood; telehealth

Introduction

The COVID-19 outbreak and the resultant reimbursement and policy changes of the Centers for Medicare and Medicaid Services have motivated a massive shift toward telehealth in routine clinical care [1]. Telehealth refers to the use of communication technologies to deliver health care remotely [2,3]. Older adults are major users of health services but may be unable or unwilling to engage in telehealth owing to the lack of internet-enabled devices, lack of internet access, or difficulties with using technology owing to sensory or functional disabilities or the lack of prior technology experience [2,4]. Older adults also face place-based challenges to telehealth as they are disproportionately likely to live in rural locations that have gaps in internet connectivity, including broadband internet [5,6].

Although the ability to engage in telehealth is affected by both individual and environmental characteristics, data constraints inhibit obtaining knowledge of how such factors jointly contribute to telehealth use. Prior studies have examined individual-level variability in predictors of telehealth use from national surveys, without regard to the neighborhood context [2,7,8]. More recent studies have examined geographic variability in telehealth use among all older adults, which have revealed, for example, smaller increases in primary care telehealth visits among rural providers early during the COVID-19 emergency compared to urban providers [1,4]. This study draws on a unique platform of linked national surveys that together afford comprehensive information on both older adults and neighborhood characteristics to assess the extent to which individual and neighborhood factors may affect engagement in telehealth among older adults with internet access.

Methods

We conducted a cross-sectional study using data of community-living older adults aged ≥65 years, who participated in the 2017 National Health and Aging Trends Study (NHATS) [9], which were linked to census tract–level data on the participants' place of residence from the American Community Survey [10]. In the 2017 NHATS, 6312 participants were interviewed: 736 participants lived in residential care facilities or nursing homes and 459 who died were excluded, yielding a final unweighted sample of 5117 participants. Measures of older adults' sociodemographic characteristics (age, gender, race, education, and income), health status (number of chronic conditions, receipt of help with daily self-care activities, and dementia status), technology use (internet use and engagement in telehealth activities), and county metropolitan status (whether participants lived in a metropolitan or nonmetropolitan county as designated by the US Office of Management and Budget) [11] were drawn from the NHATS. Census-tract measures of broadband internet subscription rates among people aged ≥65 years and a composite measure of neighborhood social deprivation [12] were derived from the American Community Survey and operationalized on the basis of the distribution within the analytic sample by tertile.

As the internet is an important technology for telehealth, we first assessed older adults' internet usage, defined by self-reports of having browsed the internet on a computer or mobile device in the last month for any reason other than emailing or texting, as previously described [7]. In this study, engagement in telehealth is defined as having browsed the internet in the last year to contact medical providers. Participants who reported browsing the internet in the past month were asked to respond with "Yes" or "No" to the following question: "in the last year have you gone on the internet or online to contact any of your medical providers…this includes making or changing medical appointments, getting test results, requesting referrals or prescriptions, or to get advice?" These activities have been referred to as "digital health technology use" [7] and "health-related internet use" [8] in other studies. We instead refer to these activities as "telehealth" because they have been an important part of the overall massive shift toward web-based health behavior during the COVID-19 emergency [13].

We then comparatively described the characteristics of older adults by internet usage across individual, census-tract, and county characteristics and assessed the adjusted odds of older adults' internet use in a multivariable logistic regression model that controls for individual, neighborhood, and county characteristics. Furthermore, among the subset of older adults who used the internet, we examined engagement in telehealth. We described the characteristics of older adults who did and those who did not engage in telehealth by individual, census-tract, and county characteristics and assessed the adjusted odds of older adults' engagement in telehealth in a multivariable logistic regression model that controls for individual, neighborhood, and county characteristics. Reported estimates are weighted with the 2017 survey weights to account for nonresponse, oversampling of subgroups (oldest-old and Black non-Hispanic persons), incomplete interviews, and replenishment of the original sample [14]. Analyses were conducted using Stata (version 13, Stata Corp), and $\alpha<.05$ indicated significance. NHATS participants provided written informed consent, and our study was approved by the institutional review board of Johns Hopkins Bloomberg School of Public Health (IRB00011232).

Results

Of an estimated 35.3 million community-living older adults, 21.1 million (59.7%) were characterized as internet users (Table 1). Relative to those who did not use the internet, those who did were younger and more likely to be male, White, more educated, higher income, in better health, and living in metropolitan counties and in census tracts at the top tertile of broadband internet subscription and the bottom tertile of social deprivation, based on unadjusted frequencies. With the exception of the number of chronic conditions and county metropolitan status, these factors were significantly associated with internet use in a fully adjusted multivariable regression model, and women had a significantly higher odds of reporting usage of the internet despite being less represented among internet users.

More than one-third (35.8%) community-living older adults who used the internet in 2017 had engaged in telehealth (Table 2). Internet users who engaged in telehealth were younger (73.5 vs 74.4 years; $P<.001$) and more likely to have a college education (54.0% vs 35.3%; $P<.001$) and be in the highest income tertile (64.4% vs 45.7%; $P<.001$) than those who did not. Internet users who engaged in telehealth were less likely to live in a nonmetropolitan county (11.5% vs 18.3%; $P=.01$) and in neighborhoods at the top 2 tertiles of social deprivation (50.4% vs 62.3%; $P<.001$), while being more likely to live in neighborhoods at the top tertile of broadband subscription (53.7% vs 38.3%).

Table 1. Internet usage among community-living older adults.

Parameters	Nonusers[a] n (weighted %)	Users[a] n (weighted %)	*P* value[b]	Adjusted odds ratio (95% CI)[c]
Older adults, row (%)	2680 (40.3)	2437 (59.7)	N/A[d]	N/A
Age (years), mean (SE)	78.4 (0.21)	74.0 (0.12)	<.001	0.90 (0.89-0.91)
Females, column (%)	1634 (57.3)	1294 (53.2)	.03	1.32 (1.11-1.58)
Non-White, column (%)	1148 (33.3)	375 (10.2)	<.001	0.31 (0.25-0.39)
Education				
High school or less, column (%)	1798 (66.9)	630 (26.0)	<.001	0.21 (0.16-0.26)
Some college, column (%)	582 (22.2)	767 (32.0)	N/A	0.48 (0.37-0.64)
College or beyond, column (%)	300 (10.9)	1040 (42.0)	N/A	Reference
Income tertiles				
<US $27,700, column (%)	1570 (55.2)	477 (18.0)	<.001	0.30 (0.23-0.39)
US $27,700-$60,000, column (%)	721 (28.8)	751 (29.7)	N/A	0.56 (0.44-0.71)
>US $60,000, column (%)	389 (16.0)	1209 (52.4)	N/A	Reference
Chronic conditions, mean (SE)[e]	2.67 (0.04)	2.01 (0.03)	<.001	0.95 (0.89-1.02)
Receiving help with daily activities[f], column (%)	638 (20.4)	203 (7.4)	<.001	0.62 (0.48-0.82)
Dementia, column (%)	528 (15.7)	58 (1.7)	<.001	0.24 (0.17-0.36)
Living in a nonmetropolitan county, column (%)	598 (21.6)	421 (15.9)	.01	0.94 (0.73-1.22)
Social deprivation index[g]				
Lowest tertile (<28), column (%)	441 (18.2)	979 (42.0)	<.001	Reference
Top 2 tertiles (>28), column (%)	2239 (81.8)	1458 (58.0)	N/A	0.73 (0.56-0.94)
Broadband subscription[h]				
Lowest tertile (<60.9%), column (%)	1403 (47.6)	618 (22.7)	<.001	0.44 (0.33-0.59)
Middle tertile (60.9%-76.0%), column (%)	769 (32.6)	802 (33.5)	N/A	0.68 (0.52-0.89)
Highest tertile (>76.0%), column (%)	508 (19.8)	1017 (43.8)	N/A	Reference

[a] Weighted estimates of the 2017 National Health and Aging Trends Study (round 7): 21.1 million nonusers and 14.2 million users.

[b] Pearson chi-square test for frequencies and the adjusted Wald test for means.

[c] Adjusted odds ratios derived from a multivariable logistic regression model that is inclusive of all measures presented in the table.

[d] N/A: not applicable.

[e] Self-reported hearing difficulty, vision difficulty, heart attack, heart disease, hypertension, stroke, hip fracture, diabetes, cancer, lung disease, and arthritis.

[f] Daily activities include eating, bathing, toileting, and dressing.

[g] 2015 social deprivation index tertile: ranges from 0 to 100; higher values indicate greater social deprivation.

[h] Broadband subscription rates among older adults in accordance with the 2017 American Community Survey (Table S2802).

Table 2. Telehealth use among community-living older adults who use the internet.

Parameters	Nonusers[a] n (weighted %)	Users[a] n (weighted %)	*P* value[b]	Adjusted odds ratio (95% CI)[c]
Older adults, row (%)	1608 (64.2)	829 (35.8)	N/A[d]	N/A
Age (years), mean (SE)	74.4 (0.14)	73.5 (0.21)	<.001	0.97 (0.95-0.98)
Females, column (%)	889 (54.6)	405 (50.6)	.12	1.10 (0.87-1.38)
Non-White, column (%)	257 (11.0)	118 (8.7)	.14	0.83 (0.58-1.21)
Education				
High school or less, column (%)	496 (31.2)	134 (16.7)	<.001	0.46 (0.34-0.63)
Some college, column (%)	528 (33.5)	239 (29.3)	N/A	0.66 (0.52-0.84)
College or beyond, column (%)	584 (35.3)	456 (54.0)	N/A	Reference
Income tertiles				
<US $27,700, column (%)	384 (22.4)	93 (10.0)	<.001	0.46 (0.33-0.64)
US $27,700-$60,000, column (%)	527 (31.9)	224 (25.6)	N/A	0.74 (0.56-0.98)
>US $60,000, column (%)	697 (45.7)	512 (64.4)	N/A	Reference
Chronic conditions, mean (SE)[e]	2.02 (0.5)	1.98 (0.06)	.68	1.06 (0.94-1.19)
Receiving help with daily activities[f], column (%)	127 (7.3)	76 (7.6)	.80	1.18 (0.76-1.82)
Dementia, column (%)	43 (1.8)	15 (1.6)	.70	1.01 (0.50-2.04)
Living in a nonmetropolitan county, column (%)	313 (18.3)	108 (11.5)	.01	0.82 (0.53-1.28)
Social deprivation index[g]				
Lowest tertile (<28), column (%)	586 (37.7)	393 (49.6)	<.001	Reference
Top 2 tertiles (>28), column (%)	1022 (62.3)	436 (50.4)	N/A	0.87 (0.67-1.13)
Broadband subscription[h]				
Lowest tertile (<60.9%), column (%)	465 (26.4)	153 (16.1)	<.001	0.61 (0.42-0.87)
Middle tertile (60.9%-76.0%), column (%)	552 (35.3)	250 (30.2)	N/A	0.77 (0.58-1.03)
Highest tertile (>76.0%), column (%)	591 (38.3)	426 (53.7)	N/A	Reference

[a] Weighted estimates of the 2017 National Health and Aging Trends Study (round 7): 13.5 million nonusers and 7.5 million users.

[b] Pearson chi-square test for frequencies and the adjusted Wald test for means.

[c] Adjusted odds ratios derived from a multivariable logistic regression model that is inclusive of all measures presented in the table.

[d] N/A: not applicable.

[e] Self-reported hearing difficulty, vision difficulty, heart attack, heart disease, hypertension, stroke, hip fracture, diabetes, cancer, lung disease, and arthritis.

[f] Daily activities include eating, bathing, toileting, and dressing.

[g] 2015 social deprivation index tertile: ranges from 0 to 100; higher values indicate greater social deprivation.

[h] Broadband subscription rates among older adults in accordance with the 2017 American Community Survey (Table S2802).

In a multivariable logistic regression model, age, education, income, and neighborhood broadband subscription remained highly associated with older adults' engagement in telehealth (Table 2): living in a neighborhood at the lowest (vs highest) tertile of broadband internet subscription was associated with a 40% lesser likelihood to engage in telehealth (adjusted odds ratio 0.61, 95% CI 0.42-0.87). No differences in telehealth use were observed by older adults' gender, race, health, neighborhood social deprivation, or county metropolitan status in the regression model.

Discussion

This nationwide study establishes that among community-living older adults who use the internet, the pervasiveness of neighborhood broadband internet is highly associated with engagement in telehealth. Prior studies have reported that more than 1 in 4 Medicare beneficiaries—more often those who are older, with lower incomes, and of Black or Hispanic racial or ethnic status—lack internet access at home [4,7] and that having more limited social and economic resources and living in a nonmetropolitan county are associated with "unreadiness" to engage in telehealth [2]. Our study extends this knowledge by drawing on a unique data source that affords the ability to differentiate older adults' internet use and engagement in telehealth alongside comprehensive information on both individual and neighborhood characteristics. Our findings indicate that, all else equal, living in a neighborhood at the lowest (vs highest) tertile of broadband internet subscription is associated with a 40% lesser likelihood to engage in telehealth. Importantly, of those characteristics that were most highly associated with engagement in telehealth (including age, education, and income), neighborhood broadband subscription stands out as uniquely amenable to policy intervention.

A recent study reported that nearly half of all fee-for-service primary care visits occurred through telehealth following the COVID-19 outbreak and that telehealth usage was disproportionately concentrated in urban and metropolitan counties [1]. Our study confirms that prior to the emergence of the COVID-19 outbreak, among internet users, living in a metropolitan county and low neighborhood social deprivation are associated with engagement in telehealth. However, there is promise in our finding that the strengths of these associations are attenuated and not significant in the context of individual and regional characteristics, including the prevalence of neighborhood broadband subscription. Although the lack of broadband internet may compound other barriers to health care access—for example, more limited availability of health care professionals and poor transportation infrastructure in rural areas [5]—prior studies of barriers to telehealth use among older adults have primarily focused on individual attitudes, capabilities, and resources [15]. While the expansion of broadband availability would remove a significant barrier to telehealth access, expansion of telehealth use will require attention to a broader set of cost and access factors such as affordability of broadband services and computer equipment, as well as education and assistance with technologies to overcome gaps in technology experience [2,6].

This cross-sectional study does not provide insight into the frequency, quality, modality, or setting of telehealth interactions. The survey information analyzed in this study was from 2017, prior to the more flexible telehealth environment promoted by Centers for Medicare & Medicaid Services since March 2020. Our study results are limited to telehealth use among the subset of older adults who reported using the internet in the previous 30 days. Audio-only telephone-based telehealth communication, which has become common during the COVID-19 emergency [16] was not measured in this study. Despite these limitations, our findings reinforce the relevance of broadband internet as a consequential factor in equitable access to telehealth and suggest possibilities from policies and programs to extend broadband internet availability and affordability [17,18]. The COVID-19 outbreak has disrupted a wide range of health-producing activities including education, health care, and commerce. In this context, study findings resonate with calls to recognize and invest in broadband internet as a social determinant of health, which is especially significant during this time of heavy reliance on remote technologies [18].

Acknowledgments

This study was supported by the National Institute on Aging (P30AG066587; T32AG000247; T32AG066576). The sponsor was not involved in the design and conduct of this analysis; the management, analysis, or interpretation of the data; the preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. JLW, SMO, and JFM had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

References

1. Bosworth A, Ruhter J, Samson LW, Sheingold S, Taplin C, Tarazi W, et al. Medicare Beneficiary Use of Telehealth Visits: Early Data from the Start of COVID-19 Pandemic. Washington, DC. Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services. 2020 Jul 28. URL: https://aspe.hhs.gov/system/files/pdf/263866/HP_IssueBrief_MedicareTelehealth_final7.29.20.pdf [accessed 2020-08-05]
2. Lam K, Lu A, Shi Y, Covinsky K. Assessing Telemedicine Unreadiness Among Older Adults in the United States During the COVID-19 Pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391. [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
3. Telehealth Programs. US Department of Health and Human Services. 2019. URL: https://www.hrsa.gov/rural-health/telehealth [accessed 2021-04-19]
4. Roberts E, Mehrotra A. Assessment of Disparities in Digital Access Among Medicare Beneficiaries and Implications for Telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389. [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
5. Henning-Smith C. The Unique Impact of COVID-19 on Older Adults in Rural Areas. J Aging Soc Policy 2020;32(4-5):396-402. [doi: 10.1080/08959420.2020.1770036] [Medline: 32475255]
6. Mossberger K, Tolbert C, Bowen D, Jimenez B. Unraveling Different Barriers to Internet Use. Urban Affairs Review 2012 Aug 13;48(6):771-810 [FREE Full text] [doi: 10.1177/1078087412453713]

7. Levine D, Lipsitz S, Linder J. Trends in Seniors' Use of Digital Health Technology in the United States, 2011-2014. JAMA 2016 Aug 02;316(5):538-540. [doi: 10.1001/jama.2016.9124] [Medline: 27483069]
8. Choi N, Dinitto DM. Internet use among older adults: association with health needs, psychological capital, and social capital. J Med Internet Res 2013 May 16;15(5):e97 [FREE Full text] [doi: 10.2196/jmir.2333] [Medline: 23681083]
9. Kasper JD, Freedman VA. National Health and Aging Trends Study User Guide: Rounds 1-9 Beta Release. Baltimore: Johns Hopkins University School of Public Health. 2020. URL: https://www.nhats.org/sites/default/files/inline-files/NHATS_User_Guide_R9_Beta_Release_6_23_2020_0.pdf [accessed 2021-05-25]
10. American Community Survey Data. United States Census Bureau. URL: https://www.census.gov/programs-surveys/acs/data.html [accessed 2020-06-15]
11. Revised Delineations of Metropolitan Statistical Areas, Micropolitan Statistical Areas, and Combined Statistical Areas, and Guidance on Uses of the Delineations of These Areas. Executive Office of the President, Office of Management and Budget, Washington, DC. 2013 Feb 28. URL: https://obamawhitehouse.archives.gov/sites/default/files/omb/bulletins/2013/b-13-01.pdf [accessed 2020-11-12]
12. Butler D, Petterson S, Phillips R, Bazemore A. Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery. Health Serv Res 2013 Apr;48(2 Pt 1):539-559 [FREE Full text] [doi: 10.1111/j.1475-6773.2012.01449.x] [Medline: 22816561]
13. 45% of adults over 65 lack online medical accounts that could help them sign up for COVID-19 vaccinations. Institute for Healthcare Policy & Innovation. 2021 Jan 15. URL: https://ihpi.umich.edu/news/45-adults-over-65-lack-online-medical-accounts-could-help-them-sign-covid-19-vaccinations [accessed 2021-04-19]
14. DeMatteis J, Freedman V, Kasper J. National Health and Aging Trends Study Development of Round 7 Survey Weights. Baltimore: Johns Hopkins University School of Public Health. 2018. URL: https://www.nhats.org/sites/default/files/2021-01/NHATS_Round_7_Weights_Technical_Paper.pdf [accessed 2021-01-25]
15. Kruse C, Fohn J, Wilson N, Nunez Patlan E, Zipp S, Mileski M. Utilization Barriers and Medical Outcomes Commensurate With the Use of Telehealth Among Older Adults: Systematic Review. JMIR Med Inform 2020 Aug 12;8(8):e20359 [FREE Full text] [doi: 10.2196/20359] [Medline: 32784177]
16. McElroy J, Day T, Becevic M. The Influence of Telehealth for Better Health Across Communities. Prev Chronic Dis 2020 Jul 16;17:E64 [FREE Full text] [doi: 10.5888/pcd17.200254] [Medline: 32678060]
17. Stauffer A, de Wit K, Read A, Kitson D. How States Are Expanding Broadband Access: New research identifies tactics for connecting unserved communities. The Pew Charitable Trusts. URL: https://www.pewtrusts.org/-/media/assets/2020/03/broadband_report0320_final.pdf [accessed 2021-05-25]
18. Benda N, Veinot T, Sieck C, Ancker J. Broadband Internet Access Is a Social Determinant of Health!. Am J Public Health 2020 Aug;110(8):1123-1125. [doi: 10.2105/AJPH.2020.305784] [Medline: 32639914]

Abbreviations

NHATS: National Health and Aging Trends Study

Edited by R Kukafka; submitted 03.12.20; peer-reviewed by E Roberts, R Zuckerman, E Brainin, T Ndabu; comments to author 16.01.21; revised version received 19.02.21; accepted 30.04.21; published 14.06.21.

Please cite as:
Okoye SM, Mulcahy JF, Fabius CD, Burgdorf JG, Wolff JL
Neighborhood Broadband and Use of Telehealth Among Older Adults: Cross-sectional Study of National Survey Data Linked With Census Data
J Med Internet Res 2021;23(6):e26242
URL: https://www.jmir.org/2021/6/e26242
doi:10.2196/26242
PMID:34125071

Review

The Use of Information and Communication Technologies by Sex Workers to Manage Occupational Health and Safety: Scoping Review

Thérèse Bernier[1], MSc; Amika Shah[1], MScCH; Lori E Ross[2], PhD; Carmen H Logie[3,4,5], PhD; Emily Seto[1], PhD

[1]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[2]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[3]Factor-Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada
[4]Women's College Research Institute, Women's College Hospital, Toronto, ON, Canada
[5]Centre for Gender and Sexual Health Equity, Vancouver, BC, Canada

Corresponding Author:
Thérèse Bernier, MSc
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
155 College St
Toronto, ON, M5T 1P8
Canada
Phone: 1 4167232089
Email: therese.bernier@mail.utoronto.ca

Abstract

Background: In many countries, sex work is criminalized, driving sex work underground and leaving sex workers vulnerable to a number of occupational health and safety risks, including violence, assault, and robbery. With the advent of widely accessible information and communication technologies (ICTs), sex workers have begun to use electronic occupational health and safety tools to mitigate these risks.

Objective: This study aims to explore the use of ICTs by sex workers for managing occupational health and safety risks and strategies for reducing these risks. This paper aims to answer the following question: what is known about sex workers' use of ICTs in the delivery of occupational health and safety strategies?

Methods: A literature review following the methodological framework for scoping reviews was conducted to analyze studies describing the use of ICTs by sex workers to mitigate occupational health and safety risks. Experimental, observational, and descriptive studies, as well as protocol papers, were included in this scoping review.

Results: Of the 2477 articles initially identified, 41 (1.66%) met the inclusion criteria. Of these studies, 71% (29/41) were published between 2015 and 2019. In these studies, the internet was the predominant ICT (24/41, 58%), followed by text messaging (10/41, 24%) and assorted communication technologies associated with mobile phones without internet access (7/41, 17%; eg, voice mail). In 56% (23/41) of the studies, sex workers located in high-income countries created occupational health and safety strategies (eg, bad date lists) and shared them through the internet. In 24% (10/41) of the studies, mostly in low- and middle-income countries, organizations external to sex work developed and sent (through text messages) occupational health and safety strategies focused on HIV. In 20% (8/41) of the studies, external organizations collaborated with the sex worker community in the development of occupational health and safety strategies communicated through ICTs; through this collaboration, concerns other than HIV (eg, mental health) emerged.

Conclusions: Although there has been an increase in the number of studies on the use of ICTs by sex workers for managing occupational health and safety over the past 5 years, knowledge of how to optimally leverage ICTs for this purpose remains scarce. Recommendations for expanding the use of ICTs by sex workers for occupational health and safety include external organizations collaborating with sex workers in the design of ICT interventions to mitigate occupational health and safety risks; to examine whether ICTs used in low- and middle-income countries would have applications in high-income countries as a substitute to the internet for sharing occupational health and safety strategies; and to explore the creation of innovative, secure, web-based communities that use existing or alternative digital technologies that could be used by sex workers to manage their occupational health and safety.

(J Med Internet Res 2021;23(6):e26085) doi:10.2196/26085

KEYWORDS

sex work; smartphone; mobile phone; occupational health and safety; online; internet; website

Introduction

Background

Sex workers encounter a number of occupational health and safety risks. In Germany [1], Thailand [2], Ireland [3], Israel [4], Cambodia [5], and Canada [6-8], studies have reported that sex workers are subjected to violence, assault, and harassment. These risks are experienced by transgender individuals who identify as women [6], cisgender women [7-9], cisgender men [10], and transgender sex workers who identify as men [10]. Although there are many categories of sex work, this paper will focus on sex work where a sex worker and their client have physical contact as commercial sex work [11].

To counteract these occupational health and safety risks, sex workers have begun to use information and communication technologies (ICTs)—most commonly, the internet and social media—to exchange tips and information [12]. In a study composed of a literature review and interviews with sex workers that was published in 2016, the Global Network of Sex Work Projects (NSWP) reported that ICTs are a new type of business tool for sex workers that has enabled a global change from outdoor sex work to indoor sex work [12]. In addition, their study reported that sex workers in high-income countries use the internet to advertise their services. In low- and middle-income countries (LMICs), sex workers provide regular clients with their phone numbers and receive calls from potential clients. In consultation with sex workers in 7 countries, ranging from low- to high-income countries, the NSWP reported that ICTs provide sex workers with the means to screen clients as well as set the terms of the encounter in advance of appointments with clients. This has led to a decrease in the rate of violence perpetrated against sex workers [12]. The predominant method for accessing ICTs was found to be through a smartphone [12]. The same study posited that the use of ICTs enhanced social cohesion among sex workers [12], which can lead to beneficial effects among sex workers in reducing occupational health and safety risks [13,14]. In addition, the NSWP highlighted in their review that the World Health Organization has published data on the value of ICTs in promoting condom use [12].

Objectives

This scoping review aims to answer the following research question: what is known about sex workers' usage of ICTs in the delivery of occupational health and safety strategies? Specifically, this study seeks to examine the types of ICTs used by sex workers, the types of occupational health and safety strategies delivered through ICTs, and the individuals and organizations engaged in the development of occupational health and safety strategies. A thorough search of the literature did not reveal any systematic reviews or scoping reviews on this topic.

Methods

Literature Review Strategy

A literature review consistent with the Preferred Reporting Items for Systematic Reviews and Meta-analyses Extension for Scoping Reviews guidelines and following the methodology pioneered by Arksey and O'Malley [15] and enhanced by Levac et al [16] was used to analyze studies on sex workers' use of ICTs in the delivery of occupational health and safety strategies.

Inclusion Criteria

Peer-reviewed studies were included if they met the following criteria: (1) published in English and (2) described an intervention in which sex workers used ICTs for occupational health and safety strategies (occupational health and safety strategies include, but are not limited to, violence prevention, having a buddy system, managing the effects of stigma on mental health, screening clients, condom use when providing sexual services, and clearly stating boundaries during the encounter), ICTs were developed for sex workers to apprise them of occupational health and safety strategies, and ICTs were developed with sex workers to exchange and communicate occupational health and safety strategies. Experimental, observational, and descriptive studies analyzing website contents, as well as protocol papers, were screened and included for review if the inclusion criteria were met. No limitations with regard to the year of publication were imposed.

Search Strategy

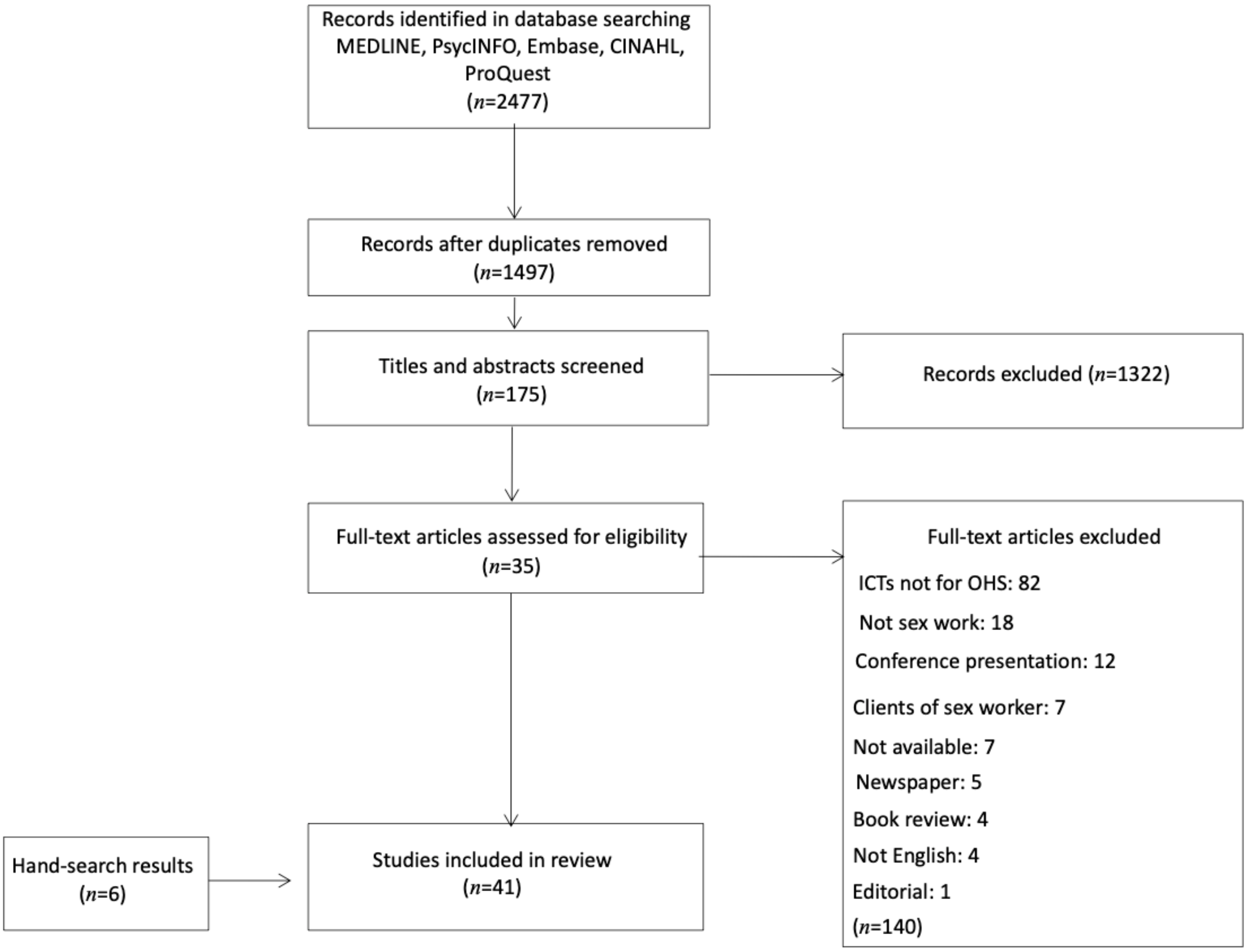
A comprehensive literature search was conducted in September 2019 using the following databases: MEDLINE, Embase, PsycINFO, CINAHL, and ProQuest Dissertations and Theses.

Search strategies were developed with the assistance of a professional librarian (Erica Lenton). Key concepts for the database search included sex worker and ICTs. The concept of occupational health and safety was screened manually because including it as a concept in the database search limited the results and excluded relevant papers. The final search strategy for MEDLINE can be found in Multimedia Appendix 1.

Selection Procedure

The selection criteria forms were developed by the primary author (TB). The initial database search resulted in 2477 articles returned (629 from MEDLINE, 713 from Embase, 420 from PsycINFO, 282 from CINAHL, and 433 from ProQuest Dissertations and Theses). The selection procedure is illustrated in the flow diagram in Figure 1. After removing 980 duplicates, 1497 articles remained. By applying the inclusion criteria, title and abstract screening was performed independently by TB and AS, resulting in 1322 excluded articles, leaving 175 articles for full-text review. The full-text review was performed independently by TB and AS. After the full-text review, 35 articles were selected for inclusion in the scoping review, and 140 were excluded. These 140 articles were excluded for the following reasons: (1) ICTs were not used for occupational health and safety strategies; (2) the article was not about sex work; (3) the article was a conference presentation (poster, oral, and plenary); (4) the article only addressed clients of sex workers and not sex workers; (5) the article was not available (eg, no longer in print); (6) the article was a newspaper article or book review; (7) the article was written in a language other than English; and (8) the article was an editorial. Disagreement between the reviewers TB and AS was resolved by discussion until a consensus was reached. A secondary search was performed by TB, consisting of screening the reference list of the included articles, which yielded 5 articles. A query to a sex work research network (Sex Work Research Hub) netted an additional single article. The search resulted in a total of 41 articles included for full data extraction and analysis. The final search strategy can be found in Multimedia Appendix 1.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart. ICT: information and communication technology; OHS: occupational health and safety.



Data Extraction

A data charting table was developed by TB, discussed with AS, and consensus was reached on the table elements that included (1) research design and if sex workers were included in the design, (2) research ethics board approval, (3) aim and goals of study, (4) type of sex work and location, (5) legislative models, (6) specific study population, (7) characteristics of the study population, (8) type of ICT, (9) occupational health and safety addressed, (10) source of occupational health and safety strategy, (11) occupational health and safety strategy, and (12) main findings. These elements were chosen to reflect the types of sex work and the associated occupational health and safety risks. TB and AS each independently charted 5 articles and then discussed the results to ensure consistency in their understanding of each table element. TB and AS then proceeded to independently complete the data extraction for all 41 articles. Any disagreement between TB and AS was resolved through discussion until a consensus was reached. At the completion of data extraction, themes were identified by examining the exposure that sex workers face with regard to occupational health and safety risks, the provenance of occupational health and safety strategies, and the ICTs used to communicate these strategies.

Results

Study Characteristics

Of the 41 studies selected for review, 7 were experimental (5 were randomized controlled trials) [17-23], 19 were observational studies [10,24-41], 13 were descriptive studies analyzing website contents [42-54], and 2 were protocol papers

[55,56]. Each paper was a unique separate study, meaning that no two papers discussed the same data.

A total of 5 studies were conducted in high-income countries (Australia, Canada, China, the United Kingdom, and the United States), and 7 studies were conducted in LMICs (Cambodia, India, Kenya, Mexico, Mozambique, South Africa, and Zimbabwe).

Study publication dates ranged from 2004 to 2019, with 71% (29/41) of the studies published between 2015 and 2019. Since 2007, according to a study published by the NSWP in 2017, ICT use by sex workers, particularly the internet and smartphones, has risen, reflecting the pace of growth in the general population [12]. For example, 33% of the Canadian adults in 2012 were using smartphones. In 2013, this percentage increased to 56% [57]. By 2018, 90% of the Canadian adults were using smartphones [58].

Main Themes

The main themes identified in the literature were related to (1) sex work context, (2) ICTs, (3) sex workers' implementation of occupational health and safety strategies through ICTs, (4) implementation of occupational health and safety strategies through ICTs by organizations external to sex workers, and (5) researchers collaborating with sex workers in the study of ICTs used for occupational health and safety strategies. Representative studies (exemplars) are presented and discussed for themes 3, 4, and 5.

Sex Work Context

Within the sex work context, four concepts were identified: occupational health and safety concerns, gender, sex work venue, and legislative models.

The 41 papers reported on one or more of the following categories of occupational health and safety concerns: (1) sexually transmitted infection and HIV prevention, testing, and risk behaviors (13 studies); (2) risk management practices (13 studies); (3) sexual and reproductive health (3 studies); (4) hepatitis B virus prevention (1 study); (5) drug use (1 study); and (6) stigma (1 study). The occupational health and safety concerns and associated studies are presented in Table 1.

Of the included studies, 41% (17/41) pertained to cisgender women [18,19,21,26-30,34,36,37,39,46,50,55,56,59]. Approximately 27% (11/41) were on cisgender men [20,31,32,40,42,43,48,51-54], and 7% (3/41) addressed the following genders: cisgender women, cisgender men, and transgender women [33,44,49]. Of the studies, 4% (2/41) addressed cisgender women, cisgender men, and transgender individuals whose gender identity was not specified [41,45], and 12% (5/41) of the studies did not specify a gender [17,22,23,25,35]. Approximately 5% (2/41) of the studies pertained to cisgender women and cisgender men [24,47], and 1 study addressed cisgender men and transgender sex workers who identified as men [10]. No studies were found that focused solely on transgender sex workers.

The sex work venue is generally regarded as indoor, outdoor, or both. Of the 13 studies that did not specify the venue, extrapolation was performed to determine the venue. Extrapolation was based on studies taking place in the same country or used information contained in the study or used other reliable sources (such as government websites). A total of 14 studies specifically mentioned indoor sex work as the type of venue in which sex workers conducted their business [25,26,31,33,34,37,43,47-52,56]. A total of 6 additional studies were extrapolated to be indoor [22-24,42,53,54]. Venues that were specified as "internet-based" [45] or "web-based" [44] were considered to be indoor sex work. Sex tourism was categorized as indoor based on the content of the study [46]. A total of 3 studies mentioned outdoor (street-based) sex work [9,17,40]; 2 were extrapolated to be outdoors [18,55]; and 8 studies addressed occupational health and safety for both indoor and outdoor (street-based) sex workers [10,27-30,35,36,59], with 5 studies extrapolated to be as such [20,21,32,39,41]. Of the 41 studies, 23 (56%) addressed indoor sex work, 5 (12%) addressed outdoor sex work, and 13 (32%) addressed sex workers in both indoor and outdoor settings.

Legislative models pertaining to sex work for each country in the study are shown in Table 2. Legislative models are categorized as (1) full criminalization, (2) partial criminalization, (3) criminalization of purchase of sex, (4) regulatory models, and (5) full decriminalization (definitions included in the notes for Table 2) [60]. Sex work is deemed to be a criminal activity in 83% (10/12) of the countries in the papers selected for this study and is regulated in 17% (2/12) of the countries in the study.

Table 1. Occupational health and safety concerns for indoor, outdoor, and indoor and outdoor sex workers.

OHS[a] addressed	Gender				
	Cisgender woman	Cisgender man	Transgender individual	All genders	Not specified
STI[b] and HIV prevention and testing					
STI prevention [25]	I&O[c]	—[d]	—	—	I
STI testing [17,36,56]	I[e], I&O	—	—	—	—
STI: Syphilis [51]	—	I	—	—	—
Regular HIV testing [17,36]	I&O	—	—	—	O[f]
HIV self-testing [18,19,56]	O, I	—	—	—	—
Reattend for HIV and STI testing [23]	—	—	—	—	I
HIV prevention [20,26,27,32,37,40,41]	I, I&O	I&O, O	—	I	—
Reduction of HIV and STI incidence [21]	I&O	—	—	—	—
ART[g] adherence [39]	I&O	—	—	—	—
Safe sex practices with clients [17,25,31,52,56]	I, I&O	I	—	—	I, O
Condomless sex [20,46,49,50,54]	I	I, I&O	I (women)	I	—
Inconsistent condom use [26,28-30]	I, I&O	—	—	—	—
Risk management practices					
Risk management [34,42,44,47]	I	I	I (women)	—	—
Personal safety [43,45]	—	I	—	I	—
Violence [10,33,59]	I, I&O	I	I (identify as men)	I	—
Violence prevention [35]	—	—	—	I&O	—
Virtual violence [34]	I	—	—	—	—
Electronic abuse [34]	I	—	—	—	—
Verbal abuse [34]	I	—	—	—	—
Harassment [59]	I&O	—	—	—	—
Harassment through verbal abuse, repeated or unwanted contact or attempts to contact through email, text, or social media [33]	—	—	—	I	—
Stalking through verbal abuse, repeated or unwanted contact or attempts to contact through email, text, or social media [33]	—	—	—	I	—
Murder [33]	—	—	—	I	—
Nonpayment [33]	—	—	—	I	—
Attempts to underpay [33]	—	—	—	I	—
Sexual and reproductive health					
Sexual and reproductive health [24,47,55,56]	O, I	I	—	—	—
Health status [46,48,53]	I	I	—	—	—
Sex worker health [45]	—	—	—	I	
HBV[h] prevention [22]	—	—	—	—	I
Drug use [51]	—	I	—	—	—
Stigma [10]	—	I&O	I&O (identify as men)	—	—

[a] OHS: occupational health and safety.

[b] STI: sexually transmitted infection.

[c] I&O: indoor and outdoor.

[d] Not available. Either no studies were found for a specific gender, or the studies found addressed all genders, or the studies did not specify a gender.

[e] I: indoor.

[f] O: outdoor.

[g] ART: antiretroviral therapy.

[h] HBV: hepatitis B virus.

Table 2. Legislative models[a,b].

Country[c]	Articles found, n (%)[d]	Full criminalization	Partial criminalization	Criminalization of purchase of sex	Regulatory models	Full decriminalization
South Africa [17][e]	1 (2)	✓				
Zimbabwe [17][e]	N/A[f]		✓ [61]			
Mozambique [17][e]	N/A		✓ [62]			
Kenya [18,19,55]	3 (7)	✓ [61]				
India [20,28-30,36,39,40,59]	8 (21)		✓			
Cambodia [37,56]	2 (5)	✓ [63]				
China [26,27]	2 (5)	✓ [64]				
Australia [22,24,42,43,49]	5 (12)				✓	
United Kingdom [23,33,34,47,50]	5 (12)		✓			
United Kingdom, Canada, and Australia [32]	1 (2)		✓ (United Kingdom)	✓ (Canada)	✓ (Australia)	
Canada [10,35,45]	3 (7)			✓		
United States [44,51-54]	5 (12)	✓				
United States and Canada [25]	1 (2)	✓ (United States)		✓ (Canada)		
United States and United Kingdom [48]	1 (2)	✓ (United States)	✓ (United Kingdom)			
Mexico [21]	1 (2)				✓	
Global [41,46]	2 (5)					
Unknown (country not specified) [31]	1 (2)					
Total	41 (100)	5 countries	4 countries	1 country	2 countries	N/A

[a] The definition of each legislative model was obtained from Platt et al [60]. *Full criminalization:* all aspects of selling and buying sex or organization of sex work are prohibited. *Partial criminalization:* organization of sex work is prohibited, including working with others, running a brothel, involvement of a third party, or soliciting. *Criminalization of purchase of sex:* often referred to as the sex-buyer model—laws penalize sex workers working together (under third-party laws), any aspect of participating in the sex trade as a third party, and buying sex. *Regulatory models:* the sale of sex is legal in licensed models or managed zones and is often accompanied by mandatory condom use, HIV and sexually transmitted infection testing, or registration. *Full decriminalization:* all aspects of adult sex work are decriminalized, but condom use is legally required in some locations (eg, New Zealand).

[b] Unless otherwise indicated, data on legislative models were obtained from Platt et al [60].

[c] A total of 12 countries in study.

[d] Percentages have been rounded.

[e] South Africa, Zimbabwe, and Mozambique were documented in the same study.

[f] N/A: not applicable.

Use of ICTs

The internet, including websites and web-based activities (defined as performing activities such as participating in a discussion forum, researching a potential client's credentials, using email, or being active on social media), accounted for 58% (24/41) of the studies. The remainder of the studies investigated the use of mobile phones without internet access; of these, 24% (10/41) examined text messaging, and the other 17% (7/41) looked at assorted communication technologies associated with mobile phones without internet access, such as phone calls, interactive voice response (IVR), audio broadcast messaging, and voice mail.

Sex Workers' Implementation of Occupational Health and Safety Strategies Through ICTs

In 56% (23/41) of the studies, occupational health and safety strategies were initiated by the sex workers themselves, using the internet (through computers or smartphones), text messaging, or mobile phones without internet access.

Using the internet, indoor sex workers were found to use an array of electronic strategies to implement risk management practices for screening clients. In the United Kingdom and Montréal (Canada), screening tools such as electronic bad-date lists (a bad-date list is a list consisting of descriptions of clients that have been abusive (eg, violence) toward sex workers) were made available to sex workers through sex worker–run organizations [33,35]. Indoor sex workers also searched for information on potential clients by consulting closed sex worker Facebook groups and web-based sex worker forums and by using Google (Alphabet Inc). Prebooking clients and negotiating services and fees through email or web chats were all part of the occupational health and safety strategies used by sex workers to screen potential clients. Through these strategies, sex workers can assess a potential client for violence, nonpayment, and other abusive behavior. These strategies were used by indoor sex workers in Vancouver (Canada) [10], the United Kingdom [33,34], and the United States [44]. Sex workers may also report abusive content to their website administrator because anonymous web-based communication can sometimes result in threats [34] or harassment [33]. In addition, by using the internet, indoor sex workers communicated safe work practices on their website advertisements. In Vancouver (Canada), sex workers advertised the use of condoms and emphasized no drugs and no alcohol during the encounter; cisgender women specified in their advertisements the services that would be provided during client encounters [45]. Sex workers in Detroit (United States) using a web-based service to advertise their services specified the services they would provide during client encounters [44]. Similar practices were used by sex workers in the United Kingdom, Florida (United States), and Sydney (Australia) through communications through web-based advertisements [47,49,52]. With the two-way electronic communication capabilities offered by the internet and with access to the internet through mobile phones, sex workers managed the risks inherent in their profession by engaging in a dialog with clients regarding their acceptable business practices, also known as safe work practices, which mitigated their occupational health and safety risks.

As another example of sex worker–initiated occupational health and safety strategies, sex workers in the United Kingdom used the text messaging app WhatsApp (Facebook, Inc) to form private chat groups to exchange information on clients [33]. In addition, in the United Kingdom, sex workers have implemented an occupational health and safety strategy known as the “buddy system”; this strategy consists of a sex worker using a mobile phone to text a colleague to apprise them about scheduled client encounters, the location of the encounters, and start and end times of the encounters [34].

Verbally communicating with clients using mobile phones is yet another sex worker–initiated occupational health and safety strategy. In India, mobile phones without internet access are the most frequently used ICT. The main occupational health and safety strategy of cisgender women was negotiating condom use with regular clients and potential clients through phone calls [28,30]. The possibility of successfully convincing the client to use a condom was dependent on the cisgender woman’s income streams. If she had no other income apart from sex work, the odds of providing services without a condom were higher [28,30].

Implementation of Occupational Health and Safety Strategies Through ICTs by Organizations External to Sex Workers

In 44% (18/41) of the studies, occupational health and safety strategies were found to be created by organizations external to sex work. These organizations were composed of researchers, public health agencies (PHAs), community-based organizations, and social enterprises working either separately or together. In the studies, the predominant occupational health and safety concern of these organizations was the transmission and prevention of sexually transmitted infection and HIV. The strategies created by these external organizations were disseminated by using the internet, text messaging, and audio broadcast messages through mobile phones. In 24% (10/41) of the studies, occupational health and safety strategies were developed without an explicit indication of sex workers’ input. The next 4 exemplar studies highlight the internet and text messaging as the ICTs used.

Using the internet, and without describing whether collaboration with sex workers occurred on occupational health and safety strategy content, external organizations have reached out to sex workers to apprise them of sexually transmitted infection and HIV prevention strategies. In the first exemplar, in the 2000s (early 21st century) in the United States, PHAs and community-based organizations initiated a health information dissemination campaign by engaging with cisgender male sex workers and their clients in web-based forums and chat rooms in 8 US cities to discuss the risks and prevention of syphilis and HIV [51]. In this study, the forum participants in New York City engaged the PHAs in discussions regarding the risks of condomless sex [51]. The next two exemplars are from China; collaboration with sex workers was not specified in either study. Two quantitative studies examined the feasibility of using the internet to communicate HIV prevention information to indoor cisgender female sex workers [26,27]. In the first study, web-based public service announcements were used to attempt to reduce inconsistent condom use among indoor cisgender female sex workers, with a small percentage of sex workers (6.7%) reporting that they were using condoms consistently after seeing the announcements [26]. In the second study, most of the sex workers surveyed (64%) indicated that they would join a web-based HIV and sexually transmitted infection prevention program [27].

The fourth exemplar features text messaging. In this study, HIV testing was the focus for outdoor cisgender female sex workers in sub-Saharan African countries. The North Star Alliance, a social enterprise, operates roadside wellness clinics in Africa. It designed and implemented a text messaging intervention for HIV prevention by formulating the content of text messages without sex worker input. A cluster randomized controlled trial with two arms was conducted, with one arm receiving the text messaging intervention (the other arm received verbal HIV prevention information at the clinic). A total of 167 sex workers were part of the intervention arm; they were located in South Africa, Zimbabwe, and Mozambique. The text messages encouraged the study participants to use condoms during sex and obtain regular HIV testing. A total of 35 messages were sent over a 29-week period. The study found that condomless sex did not diminish in the cohort using text messaging, but there was an increase in self-reported HIV testing [17].

In 20% (8/41) of the studies, organizations external to sex workers collaborated with the sex worker community in the development of occupational health and safety strategies, as demonstrated in the following four exemplars that highlight these ICTs: IVR, audio broadcast messaging, text messaging, and voice mail messaging.

In the first exemplar, a pilot study in Kolkata, India, was conducted to examine the potential of daily messaging using IVR to improve antiretroviral therapy adherence in people living with HIV. The IVR messages were developed with a community advisory board composed in part of sex workers who were part of a peer support network for sex workers. Brief messages (less than 1 min in duration) focused on strategies for self-management of three domains: medical, mental health, and nutrition and hygiene. In addition, 3 antiretroviral therapy appointment reminders were sent each month. A total of 46 participants were enrolled in the study, and 65% (30) were cisgender female sex workers. After the 1-month pilot, focus groups were convened with the participants to provide feedback on the IVR messages. The results of the focus group discussions demonstrated appreciation for the health and wellness messages. The results after 1 month indicated that medication adherence improved [39].

In the second exemplar, in an ethnographic study in Bangalore, India, audio broadcast messages were formulated by outreach workers with lived urban sex worker experience. The goal of these broadcast messages was to remind urban sex workers (cisgender women) to get tested for HIV and syphilis. In the study, the messages were sent to 230 phone numbers; 121 recipients listened to the entire message [36].

The third exemplar is a study that was conducted in Mexico at 2 locations: Tijuana, with 141 cisgender women, and Ciudad Juarez, with 129 cisgender women. Focus groups with 25 cisgender women in Tijuana provided feedback on the content and relevance of the text messages. There were two types of text messages. One type was short-term oriented, for example, encouraging cisgender women to have condoms handy before they went out for the evening. The second type was more long-term (future) oriented; for example, although condoms are expensive, the gains achieved in maintaining one's health are worth the investment. The text messages were randomly assigned to participants in both cities in the study. The goal of the study was to evaluate the impact of the short- and long-term text messages that were designed to maintain improvements in safer sex practices among drug-using and nondrug-using cisgender female sex workers. The outcome measure was the HIV and sexually transmitted infection incidence rate over the 6-month follow-up period. Cisgender women in Tijuana, where there was a lower percentage of hard drug use, benefited from future-oriented text messages that focused on the advantages of long-term safer sex maintenance practices to achieve future-oriented goals. In contrast, cisgender women in Ciudad Juarez, where there was a higher percentage of methamphetamine, cocaine, heroin (ie, hard drugs), and injectable drug use, did not benefit from either the short-term or future-oriented text messages [21].

The fourth exemplar is a pilot study conducted in Chennai, India. The initial goal of the pilot was to assess whether HIV prevention messages sent by either voice mail or text messages would be acceptable to cisgender male sex workers. Message construction was based on the results of the interviews and focus groups that were conducted using semistructured questions. In addition to acceptable message content for the HIV prevention intervention, the researchers asked the study participants if there were future additional services that they would like to receive. The cisgender men voiced their concerns regarding sexual health, negotiating condom use with clients, stigma, mental health, tuberculosis prevention, and assistance in exiting sex work [40].

The results found in the fourth exemplar were also observed in a number of studies included in this scoping review, meaning that when organizations external to sex work applied participatory methods to the research design, other issues, in addition to the prevention and transmission of sexually transmitted infection and HIV [20,21,36,39,40,56], surfaced. Specifically, sex workers voiced their concerns regarding sexual and reproductive health [40,55], stigma [40], mental health [39], tuberculosis prevention [40], and assistance in exiting sex work [40].

Researchers Surveying Sex Workers' Use of ICTs for Occupational Health and Safety Strategies

Although external organizations worked collaboratively with sex workers to develop occupational health and safety strategies, as illustrated in the aforementioned four exemplars, the use of collaborative methods such as participatory action research (PAR) and community-based participatory research (CBPR; PAR and CBPR come under the umbrella of participatory methods. Participatory methods in research emphasize research with, not on, communities and are a collaborative form of research involving the community in all aspects of the research [65,66]) were applied to studies that involved a collaboration between the sex worker community and researchers endeavoring to investigate the use of ICTs by sex workers for managing occupational health and safety. In the included studies for this scoping review, three collaborations took place in 2 high-income countries, the United Kingdom [33] and Canada [10,35], and one in India, an LMIC [59].

The next two exemplars demonstrate these collaborations. In Vancouver (Canada), a collaboration between researchers and members of the sex work community analyzed the content of qualitative interviews conducted with indoor and outdoor cisgender men and transgender individuals who identify as men to explore the safety aspects involved in transitioning from outdoor to internet-based indoor sex work, such as using webchats to assess a potential client's propensity for violence [10]. In moving to indoor sex work facilitated through the internet, the men in Vancouver (Canada) experienced less stigma perceived to be because of reduced visibility on the street [10]. However, the study found that not all sex workers in the community shifted to an indoor work environment. With more clients having followed the shift to having services provided indoors, fewer clients were available to the street-based sex workers, with an associated loss of income for these sex workers. The study found that the sex workers made up the loss of income by offering services such as anal sex that could be detrimental to their health [10].

The other exemplar was conducted in India [59]. A collaborative effort with sex workers, nongovernmental organizations, and academics resulted in the development of interview guides to study the challenges of using mobile phone communication for indoor and outdoor cisgender female sex workers. Using a mobile phone without internet access to solicit clients assisted the sex workers in moving their work indoors and avoiding the harassment associated with outdoor work [59].

In both studies, the sex workers expressed that the move to indoor work resulted in a decrease in violence, stigma, and harassment. However, the sex workers in both countries experienced a sense of loss of community because they no longer saw each other on the strolls (ie, on the streets) [10,59].

Discussion

Principal Findings

This scoping review examined the occupational health and safety risks experienced by sex workers and the use of ICTs as an occupational health and safety tool to mitigate these risks. Of the 41 studies included, 29 were published between 2015 and 2019, indicating that sex workers have begun to use a number of electronic tools to mitigate the occupational health and safety risks present in their profession. The included studies revealed that sex workers used one or more of the following ICTs: (1) the internet, (2) mobile phones, (3) messaging technology through mobile phones (IVR, voice mail, text, and audio broadcast messaging), and (4) social media apps. The findings suggest that the type of ICT used by sex workers is shaped by the country of employment. In high-income countries, the internet has become a commonly accepted commercial sex venue [67]. Moreover, internet use in high-income countries is greater than that in LMICs [68] and social media apps are commonly used. In LMICs, the predominant ICT is text messaging through mobile phones.

In 83% (10/12) of the countries included in this review, sex work is criminalized, and in 17% (2/12), it is regulated. No studies were found that addressed sex work in a country where it was decriminalized (in New Zealand, where decriminalization has occurred, sex workers are protected by labor laws that enhance their occupational health and safety, and mobile phones are considered an essential safety tool [69]). Legislative models that criminalize sex work in the country of employment were just one of the structural determinants that influenced a sex worker's occupational health and safety risks [70]. Other structural determinants such as working conditions, location, and income outside of sex work [28,30,71] all affected whether a condom was used during the provision of services and the type of services provided [10,50]. In public health studies, condom use is the recommended method for preventing the acquisition and transmission of sexually transmitted infections and HIV, and regular and frequent testing is also recommended to detect possible acquisition. In LMICs, the occupational health and safety strategies delivered through ICTs found in the included studies reflected this public health approach. However, these studies did not address possible financial and societal barriers that could impede sex workers—particularly women—from accessing HIV testing [72].

The criminalization of sex work and the accompanying stigma, as well as the technology available in high-income countries and LMICs, directed the choice of ICTs used by sex workers. In collaborative research projects in LMICs, discretion in receiving occupational health and safety risk mitigation strategies was the overriding concern; therefore, IVR, voice mail, and audio broadcast messaging were the preferred communication methods, as opposed to a text message that could inadvertently be seen by a family member on a sex worker's mobile phone [36,39,40]. In high-income countries, the internet is the prevalent ICT. However, the use of the internet has been impeded by the passage of the 2018 Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA) and Stop Enabling Sex Traffickers Act (SESTA) in the United States, which prohibit sex work content on websites, including harm reduction techniques and advertising [73-75]. Backpage [76] was a popular site for sex worker advertisements [44], but after the passage of the acts, its management no longer allowed sex workers to advertise on the site. The FOSTA-SESTA have also affected Canadian sex workers because 60% of the sex workers used to advertise on Backpage [76] before it was banned [77]. Craigslist, a popular web-based classified advertisements site in Toronto (Canada) and environs, has also banned all sex work–related advertisements [78]. Sex worker organizations are concerned that the safety precautions that internet-based sex workers were able to take are now disappearing because sex workers are being pushed back to outdoor work [73,77]. A recommended area of research would be to examine whether the use of text messaging, IVR, voice mail, and audio broadcast messaging could serve as safe alternative electronic spaces in high-income countries to replace those that have disappeared, not only for advertising purposes but also for sex workers to be able to once more create secure online communities to manage the structural determinants that constrain their health and safety. However, the protection of information communicated through these ICTs would depend on the laws governing electronic communication in these countries (in Canada, the Supreme Court has ruled that electronic conversations are protected by that country's privacy laws [79]). A further recommendation is to examine the creation of new secure online communities using existing or alternative digital technologies, for example, Red Cloud [80] in Australia, a web hosting enterprise that does not come under the FOSTA-SESTA.

This review found that the predominant occupational health and safety concern of organizations external to sex work was the transmission and prevention of sexually transmitted infection and HIV, and the ICT interventions developed by these organizations consisted of pushing sexually transmitted infection and HIV prevention and testing messaging out to either the internet or to sex workers' mobile phones. Only a small number of studies included in this scoping review used collaborative research methods such as PAR and CBPR to either design or study an ICT intervention. In these collaborative studies, sexual and reproductive health, stigma and its associated mental health effects, and violence emerged as the occupational health and safety risks that sex workers were most concerned about [10,33,40,56,59]. PAR and CBPR fall under the umbrella of participatory methods. When sex workers are involved in the research project as members of the research team, the research is no longer a passive exercise [81]. The sex workers' agency is leveraged, and their experiences, both personal and professional, are discussed. Through participatory research methodology used in previous studies with sex workers, structural determinants that affect sex workers' occupational health and safety, such as violence [82], stigma, social norms, lack of support from family and friends [82,83], and social support from community resources and sex worker colleagues emerged, as well as their resilience skills in managing their circumstances [83]. On the basis of these findings, it is recommended that future research involve sex workers in the formulation of the research question and scope of the study to accurately identify the issues that sex workers face in their personal and professional lives. A further area of research would be to compare the longevity and continued effectiveness of interventions designed with sex workers with those designed without sex worker input.

Limitations and Strengths

A limitation of this review is that, for feasibility reasons, only studies published in English were included. By limiting the studies to English, relevant findings may have been missed. In addition, the gray literature was not consulted, which may have resulted in novel ways of using ICTs by sex workers not being represented. Another limitation is that no studies were found that focused solely on transgender sex workers, and studies that did include transgender individuals combined them with cisgender women and cisgender men.

The strengths of this review include the rigor of the methods, including the use of duplication at each phase of the review (article screening and selection, data extraction, and full-text analysis), the absence of limitations with respect to year of publication, and the inclusion of a wide range of study designs. In addition, this study provides a broad overview of the different aspects of ICT use among sex workers across the globe because studies published during the last 15 years were included to account for recent developments in electronic communications technology.

Conclusions

This study provides important insights into how the use of ICTs during the first two decades of the 21st century has shaped the methods that sex workers use to keep themselves safe on the job in environments where sex work is criminalized. This review found that a considerable amount of research on ICT use demonstrated that internet-based sex workers, that is, sex workers using the internet to conduct business, were active creators and sharers of the occupational health and safety strategies used in their work. This review also found that researchers, PHAs, and social enterprises use text messaging as their ICT of choice for pushing sexually transmitted infection and HIV prevention and testing messages out to sex workers' mobile phones, although this is not the main occupational health and safety concern for sex workers. Only a small number of studies in this review engaged sex workers in collaborative research; yet, these collaborations yielded insights into the lives of sex workers and provided direction for future research efforts. Applying participatory methods that engage sex workers of all genders as active participants in the research process will produce applications of digital technology geared toward their occupational health and safety needs.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE search strategy.
[DOCX File , 15 KB - jmir_v23i6e26085_app1.docx]

References

1. Groneberg DA, Molliné M, Kusma B. Sex work during the World Cup in Germany. Lancet 2006 Sep 02;368(9538):840-841. [doi: 10.1016/S0140-6736(06)69324-0] [Medline: 16950356]
2. Sanders E. Researching prostitution: the methodological challenges of researching sex work in Thailand. In: Proceedings of the UK Postgraduate Conference in Gender Studies. 2006 Presented at: UK Postgraduate Conference in Gender Studies; June 21-22, 2006; University of Leeds, UK p. 1-13 URL: https://gender-studies.leeds.ac.uk/wp-content/uploads/sites/53/2013/02/epaper6-erin-sanders.pdf
3. Beegan R, Moran J. Prostitution and sex work: situating Ireland's new law on prostitution in the radical and liberal feminist paradigms. Irish J Appl Soc Stud 2017;17(1):6 [FREE Full text] [doi: 10.21427/D7BQ6T]
4. Cwikel J, Ilan K, Chudakov B. Women brothel workers and occupational health risks. J Epidemiol Community Health 2003 Oct;57(10):809-815 [FREE Full text] [doi: 10.1136/jech.57.10.809] [Medline: 14573588]

5. Page K, Stein E, Sansothy N, Evans J, Couture M, Sichan K, Young Women's Health Study Collaborative*. Sex work and HIV in Cambodia: trajectories of risk and disease in two cohorts of high-risk young women in Phnom Penh, Cambodia. BMJ Open 2013 Sep 10;3(9):e003095 [FREE Full text] [doi: 10.1136/bmjopen-2013-003095] [Medline: 24022389]
6. Page MM. Fighting for Homewood: Gentrification and the history of violent struggle over trans sex workers? strolls in Canada. In: Durisin EM, editor. Red Light Labour: Sex Work Regulation, Agency, and Resistance. Vancouver, BC: UBC Press; 2018:272-280.
7. Benoit C, McCarthy B, Jansson M. Stigma, sex work, and substance use: a comparative analysis. Sociol Health Illn 2015 Mar;37(3):437-451. [doi: 10.1111/1467-9566.12201] [Medline: 25688450]
8. Silliker A. Sex workers facing increased safety risks. Canadian Occupational Safety. 2017. URL: https://www.thesafetymag.com/ca/topics/legislation/sex-workers-facing-increased-safety-risks/185983 [accessed 2021-06-10]
9. Lazarus L, Deering KN, Nabess R, Gibson K, Tyndall MW, Shannon K. Occupational stigma as a primary barrier to health care for street-based sex workers in Canada. Cult Health Sex 2012;14(2):139-150 [FREE Full text] [doi: 10.1080/13691058.2011.628411] [Medline: 22084992]
10. Argento E, Taylor M, Jollimore J, Taylor C, Jennex J, Krusi A, et al. The loss of boystown and transition to online sex work: strategies and barriers to increase safety among men sex workers and clients of men. Am J Mens Health 2018 Nov 28;12(6):1994-2005 [FREE Full text] [doi: 10.1177/1557988316655785] [Medline: 27352925]
11. Overs C. Sex workers: part of the solution. World Health Organization. 2002. URL: https://www.who.int/hiv/topics/vct/sw_toolkit/115solution.pdf [accessed 2021-06-10]
12. Smart service provider's guide to ICT and sex work. Global Network of Sex Work Projects. 2017. URL: http://www.nswp.org/resource/smart-service-providers-guide-ict-and-sex-work [accessed 2018-09-24]
13. Kerrigan D, Telles P, Torres H, Overs C, Castle C. Community development and HIV/STI-related vulnerability among female sex workers in Rio de Janeiro, Brazil. Health Educ Res 2008 Mar;23(1):137-145. [doi: 10.1093/her/cym011] [Medline: 17363361]
14. Argento E, Duff P, Bingham B, Chapman J, Nguyen P, Strathdee SA, et al. Social cohesion among sex workers and client condom refusal in a Canadian setting: implications for structural and community-led interventions. AIDS Behav 2016 Jun;20(6):1275-1283 [FREE Full text] [doi: 10.1007/s10461-015-1230-8] [Medline: 26499335]
15. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
16. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
17. Govender K, Beckett S, Masebo W, Braga C, Zambezi P, Manhique M, et al. Effects of a Short Message Service (SMS) intervention on reduction of HIV risk behaviours and improving HIV testing rates among populations located near roadside wellness clinics: a cluster randomised controlled trial in South Africa, Zimbabwe and Mozambique. AIDS Behav 2019 Nov;23(11):3119-3128 [FREE Full text] [doi: 10.1007/s10461-019-02427-6] [Medline: 30771133]
18. Kelvin EA, George G, Mwai E, Kinyanjui S, Romo ML, Odhiambo JO, et al. A randomized controlled trial to increase HIV testing demand among female sex workers in Kenya through announcing the availability of HIV self-testing via text message. AIDS Behav 2019 Jan 14;23(1):116-125 [FREE Full text] [doi: 10.1007/s10461-018-2248-5] [Medline: 30109456]
19. George G, Chetty T, Strauss M, Inoti S, Kinyanjui S, Mwai E, et al. Costing analysis of an SMS-based intervention to promote HIV self-testing amongst truckers and sex workers in Kenya. PLoS One 2018;13(7):e0197305 [FREE Full text] [doi: 10.1371/journal.pone.0197305] [Medline: 29979704]
20. Mimiaga MJ, Thomas B, Biello K, Johnson BE, Swaminathan S, Navakodi P, et al. A pilot randomized controlled trial of an integrated in-person and mobile phone delivered counseling and text messaging intervention to reduce HIV transmission risk among male sex workers in Chennai, India. AIDS Behav 2017 Nov;21(11):3172-3181. [doi: 10.1007/s10461-017-1884-5] [Medline: 28831618]
21. Patterson TL, Semple SJ, Abramovitz D, Harvey-Vera A, Pines HA, Verdugo S, et al. Impact of time perspectives on texting intervention to reduce HIV/STI transmission among female sex workers in Tijuana and Ciudad Juarez, Mexico. J Behav Med 2019 Feb;42(1):111-127 [FREE Full text] [doi: 10.1007/s10865-018-9948-1] [Medline: 29987740]
22. McIver R, Dyda A, McNulty AM, Knight V, Wand HC, Guy RJ. Text message reminders do not improve hepatitis B vaccination rates in an Australian sexual health setting. J Am Med Inform Assoc 2016 Apr;23(e1):88-92 [FREE Full text] [doi: 10.1093/jamia/ocv145] [Medline: 26499103]
23. Nyatsanza F, McSorley J, Murphy S, Brook G. 'It's all in the message': the utility of personalised short message service (SMS) texts to remind patients at higher risk of STIs and HIV to reattend for testing-a repeat before and after study. Sex Transm Infect 2016 Dec;92(5):393-395. [doi: 10.1136/sextrans-2015-052216] [Medline: 26670912]
24. Varma R, Chung C, Townsend A, Power M. Sexual health-related information delivery - are patient information leaflets still relevant? Sex Health 2016 Jun;13(3):289-291. [doi: 10.1071/SH15205] [Medline: 26909680]
25. Cunningham S, Kendall TD. Risk behaviours among internet-facilitated sex workers: evidence from two new datasets. Sex Transm Infect 2010 Dec;86 Suppl 3:100-105. [doi: 10.1136/sti.2010.044875] [Medline: 20852309]
26. Xiao Z, Li X, Lin D, Tam CC. Mass media and HIV/AIDS prevention among female sex workers in Beijing, China. J Health Commun 2015;20(9):1095-1106. [doi: 10.1080/10810730.2015.1018575] [Medline: 25950448]

27. Hong Y, Li X, Fang X, Lin X, Zhang C. Internet use among female sex workers in China: implications for HIV/STI prevention. AIDS Behav 2011 Feb;15(2):273-282. [doi: 10.1007/s10461-010-9846-1] [Medline: 21082341]
28. Navani-Vazirani S, Solomon D, Gopalakrishnan, Heylen E, Srikrishnan AK, Vasudevan CK, et al. Mobile phones and sex work in South India: the emerging role of mobile phones in condom use by female sex workers in two Indian states. Cult Health Sex 2015;17(2):252-265 [FREE Full text] [doi: 10.1080/13691058.2014.960002] [Medline: 25301669]
29. Mahapatra B. HIV risk behaviors among female sex workers using cell phone for client solicitation in India. J AIDS Clinic Res 2012;01(S1):1-6 [FREE Full text] [doi: 10.4172/2155-6113.s1-014]
30. S N. The role of sex work pay in moderating the effect of mobile phone solicitation on condom practices: an analysis of female sex workers in India. HSOA J AIDS Clin Res STDs 2017 Jun 16;4(1):1-8. [doi: 10.24966/acrs-7370/100008]
31. Parsons JT, Koken JA, Bimbi DS. The use of the internet by gay and bisexual male escorts: sex workers as sex educators. AIDS Care 2004 Nov;16(8):1021-1035. [doi: 10.1080/09540120412331292405] [Medline: 15511734]
32. MacGibbon J, Minichiello V, Prestage G, Bell S, Cox C, Donovan B, et al. How male sex workers and their clients shifted from reluctance about HIV pre-exposure prophylaxis to advocating for its use: a longitudinal mixed methods study. AIDS Behav 2020 Mar;24(3):782-790. [doi: 10.1007/s10461-019-02618-1] [Medline: 31367966]
33. Campbell R, Sanders T, Scoular J, Pitcher J, Cunningham S. Risking safety and rights: online sex work, crimes and 'blended safety repertoires'. Br J Sociol 2019 Sep;70(4):1539-1560. [doi: 10.1111/1468-4446.12493] [Medline: 30318604]
34. Davies K, Evans L. A virtual view of managing violence among British escorts. Deviant Behavior 2007 Sep 10;28(6):525-551. [doi: 10.1080/01639620701316830]
35. Strohmayer A, Clamen J, Laing M. Technologies for social justice: lessons from sex workers on the front lines. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow Scotland UK p. 1-14. [doi: 10.1145/3290605.3300882]
36. Sambasivan N. Production of use: reconceptualizing the user in low-income communities in urban India. Doctoral Dissertation, University of California, Irvine. 2012. URL: https://isr.uci.edu/content/production-use-re-conceptualizing-user-low-income-communities-urban-india [accessed 2021-06-11]
37. Brody C, Dhaliwal S, Tuot S, Johnson M, Pal K, Yi S. Are text messages a feasible and acceptable way to reach female entertainment workers in Cambodia with health messages? A cross-sectional phone survey. JMIR Mhealth Uhealth 2016 May 20;4(2):e52 [FREE Full text] [doi: 10.2196/mhealth.5297] [Medline: 27207374]
38. Panchanadeswaran S, Unnithan AM, Chacko S, Brazda M, Kuruppu S. What’s technology got to do with it? Exploring the impact of mobile phones on female sex workers’ lives and livelihood in India. Gender Technol Develop 2017 Oct 23;21(1-2):152-167. [doi: 10.1080/09718524.2017.1385318]
39. Swendeman D, Jana S, Ray P, Mindry D, Das M, Bhakta B. Development and pilot testing of daily Interactive Voice Response (IVR) calls to support antiretroviral adherence in India: a mixed-methods pilot study. AIDS Behav 2015 Jun;19 Suppl 2:142-155 [FREE Full text] [doi: 10.1007/s10461-014-0983-9] [Medline: 25638037]
40. Thomas B, Closson EF, Biello K, Menon S, Navakodi P, Dhanalakshmi A, et al. Development and open pilot trial of an HIV-prevention intervention integrating mobile-phone technology for male sex workers in Chennai, India. Arch Sex Behav 2017 May;46(4):1035-1046 [FREE Full text] [doi: 10.1007/s10508-015-0665-3] [Medline: 26714684]
41. Rassool GH. Current issues and forthcoming events. Wiley Online Library 2003;43(3):225-229. [doi: 10.1046/j.1365-2648.2003.02701.x-i3]
42. MacPhail C, Scott J, Minichiello V. Technology, normalisation and male sex work. Cult Health Sex 2015;17(4):483-495. [doi: 10.1080/13691058.2014.951396] [Medline: 25215634]
43. McLean A. ‘You can do it from your sofa’: the increasing popularity of the internet as a working site among male sex workers in Melbourne. J Sociol 2013 Mar 08;51(4):887-902. [doi: 10.1177/1440783313480416]
44. Moorman JD, Harrison K. Gender, race, and risk: intersectional risk management in the sale of sex online. J Sex Res 2016 Sep;53(7):816-824. [doi: 10.1080/00224499.2015.1065950] [Medline: 26488687]
45. Kille J, Bungay V, Oliffe J, Atchison C. A content analysis of health and safety communications among internet-based sex work advertisements: important information for public health. J Med Internet Res 2017 Apr 13;19(4):e111 [FREE Full text] [doi: 10.2196/jmir.6746] [Medline: 28408364]
46. Gezinski LB, Karandikar S, Levitt A, Ghaffarian R. 'Total girlfriend experience': examining marketplace mythologies on sex tourism websites. Cult Health Sex 2016 Jul 13;18(7):785-798. [doi: 10.1080/13691058.2015.1124457] [Medline: 26760690]
47. Burghart KO. What’s on sale? A discourse analysis of four distinctive online escort advertisement websites. Sexual Cult 2017 Nov 13;22(1):316-335. [doi: 10.1007/s12119-017-9469-z]
48. Lee-Gonyea JA, Castle T, Gonyea NE. Laid to order: male escorts advertising on the internet. Deviant Behav 2009 Apr 03;30(4):321-348. [doi: 10.1080/01639620802168858]
49. Blackledge E, Thng C, McIver R, McNulty A. Rates of advertised condomless sex in the online profiles of private sex workers: a cross-sectional study. Sex Health 2018 Feb;15(1):86-88. [doi: 10.1071/SH17068] [Medline: 28641708]
50. Eccles C, Clarke J. Levels of advertised unprotected vaginal and oral sex by independent indoor female sex workers in West Yorkshire, UK. Sex Transm Infect 2014 Mar;90(1):36-37. [doi: 10.1136/sextrans-2013-051150] [Medline: 24078374]

51. Taylor M, Montoya JA, Cantrell R, Mitchell SJ, Williams M, Jordahl L, et al. Interventions in the commercial sex industry during the rise in syphilis rates among men who have sex with men (MSM). Sex Transm Dis 2005 Oct;32(10 Suppl):53-59 [FREE Full text] [doi: 10.1097/01.olq.0000180453.31255.2d] [Medline: 16205294]
52. Blackwell CW, Dziegielewski SF. Risk for a price: sexual activity solicitations in online male sex worker profiles. J Soc Serv Res 2013 Mar;39(2):159-170. [doi: 10.1080/01488376.2012.744617]
53. Phua VC, Ciambrone D, Vazquez O. Advertising health status in male sex workers' online ads. J Men Stud 2009 Nov 5;17(3):251-258. [doi: 10.3149/jms.1703.251]
54. Pruitt MV. Online boys: male-for-male internet escorts. Sociol Focus 2005 Aug;38(3):189-203. [doi: 10.1080/00380237.2005.10571265]
55. Ampt FH, Mudogo C, Gichangi P, Lim MSC, Manguro G, Chersich M, et al. WHISPER or SHOUT study: protocol of a cluster-randomised controlled trial assessing mHealth sexual reproductive health and nutrition interventions among female sex workers in Mombasa, Kenya. BMJ Open 2017 Aug 18;7(8):e017388 [FREE Full text] [doi: 10.1136/bmjopen-2017-017388] [Medline: 28821530]
56. Brody C, Tuot S, Chhoun P, Swendeman D, Kaplan KC, Yi S. Mobile Link - a theory-based messaging intervention for improving sexual and reproductive health of female entertainment workers in Cambodia: study protocol of a randomized controlled trial. Trials 2018 Apr 19;19(1):235 [FREE Full text] [doi: 10.1186/s13063-018-2614-7] [Medline: 29673381]
57. The Canadian Press. Smartphone use way up in Canada, Google finds. CBC/Radio-Canada. 2013. URL: https://www.cbc.ca/news/business/smartphone-use-way-up-in-canada-google-finds-1.1384916 [accessed 2020-10-14]
58. Smartphone use and smartphone habits by gender and age group. Statistics Canada. 2019. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210011501 [accessed 2020-10-14]
59. Panchanadeswaran S, Unnithan AM, Chacko S, Brazda M, Kuruppu S. What’s technology got to do with it? Exploring the impact of mobile phones on female sex workers’ lives and livelihood in India. Gender Technol Develop 2017 Oct 23;21(1-2):152-167. [doi: 10.1080/09718524.2017.1385318]
60. Platt L, Grenfell P, Meiksin R, Elmes J, Sherman SG, Sanders T, et al. Associations between sex work laws and sex workers' health: a systematic review and meta-analysis of quantitative and qualitative studies. PLoS Med 2018 Dec 11;15(12):e1002680 [FREE Full text] [doi: 10.1371/journal.pmed.1002680] [Medline: 30532209]
61. Sexuality, poverty and law programme. Institute of Development Studies. URL: http://spl.ids.ac.uk/sexworklaw/countries [accessed 2020-07-23]
62. Tiyane Vavasate, Pathfinder International, and Vrije Universiteit, Amsterdam. 2016. URL: https://www.pathfinder.org/wp-content/uploads/2019/01/Sex-Work-and-Violence-in-Mozambique.pdf [accessed 2021-06-11]
63. Law on suppression of human trafficking and sexual exploitation. Kingdom of Cambodia. 2008. URL: https://www2.ohchr.org/english/bodies/crc/docs/CRC.C.OPSC.KHM.1.pdf [accessed 2021-06-11]
64. Ling B. Prostitution and female trafficking in China. OpenEdition J 2018 Jun 01;2018(1-2):65-74 [FREE Full text] [doi: 10.4000/chinaperspectives.7742]
65. Green J, Thorogood N. Qualitative Methods for Health Research. 3rd Ed. London, United Kingdom: SAGE Publications Ltd; 2013:1-360.
66. Flicker S, Savan B, Kolenda B, Mildenberger M. A snapshot of community-based research in Canada: Who? What? Why? How? Health Educ Res 2008 Mar;23(1):106-114. [doi: 10.1093/her/cym007] [Medline: 17322572]
67. More bang for your buck. The Economist. 2014 Aug 7. URL: https://www.economist.com/news/briefing/21611074-how-new-technology-shaking-up-oldest-business-more-bang-your-buck [accessed 2020-04-01]
68. Kraidy M. The internet as a mass communication medium. Journalism and Mass Communication – Vol. II. 2011. URL: https://www.eolss.net/sample-chapters/C04/E6-33-03.pdf [accessed 2021-06-11]
69. Hill B. A guide to occupational health and safety in the New Zealand sex industry.: Occupational Safety and Health Service; 2004. URL: http://espu-usa.com/espu-ca/wp-content/uploads/2008/02/nz-health-and-safety-handbook.pdf
70. Sex workers, HIV and AIDS. avert.org. 2019. URL: https://www.avert.org/professionals/hiv-social-issues/key-affected-populations/sex-workers [accessed 2021-06-11]
71. Smith MD, Seal DW. Motivational influences on the safer sex behavior of agency-based male sex workers. Arch Sex Behav 2008 Oct 21;37(5):845-853 [FREE Full text] [doi: 10.1007/s10508-008-9341-1] [Medline: 18288599]
72. Tokar A, Broerse JE, Blanchard J, Roura M. HIV testing and counseling among female sex workers: a systematic literature review. AIDS Behav 2018 Aug;22(8):2435-2457 [FREE Full text] [doi: 10.1007/s10461-018-2043-3] [Medline: 29464430]
73. Blunt D, Wolf A. Erased: the impact of FOSTA-SESTA and the removal of backpage on sex workers. Anti-Traffic Rev 2020 Apr 27(14):117-121. [doi: 10.14197/atr.201220148]
74. Markowicz K. Congress' awful anti-sex-trafficking law has only put sex workers in danger and wasted taxpayer money. Insider. 2019. URL: https://www.businessinsider.com/fosta-sesta-anti-sex-trafficking-law-has-been-failure-opinion-2019-7 [accessed 2021-06-11]
75. Soderberg-Rivkin D. The lessons of FOSTA-SESTA from a former content moderator. Medium. 2020. URL: https://medium.com/@Daisy_Soderberg_Rivkin/the-lessons-of-fosta-sesta-from-a-former-content-moderator-24ab256dc9e5 [accessed 2020-07-29]
76. Backpage. URL: http://backpage.com/ [accessed 2021-06-16]

77. Porth K. SESTA/FOSTA: Censoring sex workers from websites sets a dangerous precedent. Pivot Legal Society. 2018. URL: http://www.pivotlegal.org/sesta_fosta_censoring_sex_workers_from_websites_sets_a_dangerous_precedent [accessed 2019-04-29]
78. Sage J. Cyborgology: the impact of FOSTA/SESTA on online sex work communities. The Society Pages. 2018. URL: https://thesocietypages.org/cyborgology/2018/04/02/the-impact-of-fosta-sesta-on-online-sex-work-communities/ [accessed 2019-04-29]
79. Wilson C. Privacy expectations in text messages - can police look at my texts?. 2019. URL: https://wilsoncriminaldefence.com/can-police-look-at-my-texts/ [accessed 2021-06-11]
80. Red Cloud. URL: https://redcloudhosting.com.au/ [accessed 2021-06-16]
81. van der Meulen E. Action research with sex workers: dismantling barriers and building bridges. Act Res 2011 Jul 20;9(4):370-384 [FREE Full text] [doi: 10.1177/1476750311409767]
82. Kloek M, Dijkstra M. Sex work, stigma and violence in the Netherlands. SOAids Nederland - Aidsfonds. 2018. URL: http://www.nswp.org/sites/nswp.org/files/sex_work_stigma_and_violence_in_the_netherlands_reportdigital.pdf [accessed 2021-06-11]
83. Burnes TR, Rojas EM, Delgado I, Watkins TE. "Wear Some Thick Socks If You Walk in My Shoes": Agency, resilience, and well-being in communities of North American sex workers. Arch Sex Behav 2018 Jul;47(5):1541-1550. [doi: 10.1007/s10508-016-0915-z] [Medline: 28280948]

Abbreviations

CBPR: community-based participatory research
FOSTA: Fight Online Sex Trafficking Act
ICT: information and communication technology
IVR: interactive voice response
LMIC: low- and middle-income country
NSWP: Global Network of Sex Work Projects
PAR: participatory action research
PHA: public health agency
SESTA: Stop Enabling Sex Traffickers Act

Edited by R Kukafka, G Eysenbach; submitted 26.11.20; peer-reviewed by C Mudogo, V Minichiello; comments to author 29.01.21; revised version received 17.02.21; accepted 30.05.21; published 24.06.21.

Please cite as:
Bernier T, Shah A, Ross LE, Logie CH, Seto E
The Use of Information and Communication Technologies by Sex Workers to Manage Occupational Health and Safety: Scoping Review
J Med Internet Res 2021;23(6):e26085
URL: https://www.jmir.org/2021/6/e26085/
doi:10.2196/26085
PMID:34185001

Original Paper

Deep Learning Application for Vocal Fold Disease Prediction Through Voice Recognition: Preliminary Development Study

Hao-Chun Hu[1,2,3], MD; Shyue-Yih Chang[4], MD; Chuen-Heng Wang[5], MSc; Kai-Jun Li[2], MSc; Hsiao-Yun Cho[2,6], MD; Yi-Ting Chen[5], MSc; Chang-Jung Lu[4], MSc; Tzu-Pei Tsai[4], MSc; Oscar Kuang-Sheng Lee[1,7,8,9], MD, PhD

[1]Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[2]Department of Otorhinolaryngology-Head and Neck Surgery, Fu Jen Catholic University Hospital, Fu Jen Catholic University, New Taipei City, Taiwan

[3]School of Medicine, College of Medicine, Fu Jen Catholic University, New Taipei City, Taiwan

[4]Voice Center, Department of Otolaryngology, Cheng Hsin General Hospital, Taipei, Taiwan

[5]Muen Biomedical and Optoelectronic Technologist Inc, Taipei, Taiwan

[6]Graduate Institute of Business Administration, Fu Jen Catholic University, New Taipei City, Taiwan

[7]Department of Orthopedics, China Medical University Hospital, Taichung, Taiwan

[8]Stem Cell Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan

[9]Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan

Corresponding Author:
Oscar Kuang-Sheng Lee, MD, PhD
Institute of Clinical Medicine
National Yang Ming Chiao Tung University
No 155, Section 2, Li-Nong Street, Beitou District
Taipei, 11221
Taiwan
Phone: 886 2 28757391
Fax: 886 2 28757841
Email: oscarlee9203@gmail.com

Abstract

Background: Dysphonia influences the quality of life by interfering with communication. However, a laryngoscopic examination is expensive and not readily accessible in primary care units. Experienced laryngologists are required to achieve an accurate diagnosis.

Objective: This study sought to detect various vocal fold diseases through pathological voice recognition using artificial intelligence.

Methods: We collected 189 normal voice samples and 552 samples of individuals with voice disorders, including vocal atrophy (n=224), unilateral vocal paralysis (n=50), organic vocal fold lesions (n=248), and adductor spasmodic dysphonia (n=30). The 741 samples were divided into 2 sets: 593 samples as the training set and 148 samples as the testing set. A convolutional neural network approach was applied to train the model, and findings were compared with those of human specialists.

Results: The convolutional neural network model achieved a sensitivity of 0.66, a specificity of 0.91, and an overall accuracy of 66.9% for distinguishing normal voice, vocal atrophy, unilateral vocal paralysis, organic vocal fold lesions, and adductor spasmodic dysphonia. Compared with the accuracy of human specialists, the overall accuracy rates were 60.1% and 56.1% for the 2 laryngologists and 51.4% and 43.2% for the 2 general ear, nose, and throat doctors.

Conclusions: Voice alone could be used for common vocal fold disease recognition through a deep learning approach after training with our Mandarin pathological voice database. This approach involving artificial intelligence could be clinically useful for screening general vocal fold disease using the voice. The approach includes a quick survey and a general health examination. It can be applied during telemedicine in areas with primary care units lacking laryngoscopic abilities. It could support physicians when prescreening cases by allowing for invasive examinations to be performed only for cases involving problems with automatic recognition or listening and for professional analyses of other clinical examination results that reveal doubts about the presence of pathologies.

(J Med Internet Res 2021;23(6):e25247) doi:10.2196/25247

KEYWORDS

artificial intelligence; convolutional neural network; dysphonia; pathological voice; vocal fold disease; voice pathology identification

Introduction

The impact of a voice disorder has been increasingly recognized as a public health concern. Dysphonia influences the quality of physical, social, and occupational aspects of life by interfering with communication [1]. A nationwide insurance claims data analysis of treatment seeking for dysphonia showed a prevalence rate of 0.98% among 55 million individuals [2], and this rate reached 2.5% among those older than 70 years [2]. However, the overall dysphonia incidence for the aging population is estimated to be much higher (12%-35%) [3], which may imply that dysphonia is commonly overlooked by patients, resulting in underdiagnosis.

According to the state-of-the-art clinical practice guidelines for dysphonia of the American Academy of Otolaryngology-Head and Neck Surgery Foundation, a laryngoscopic examination is recommended if dysphonia fails to resolve or improve within 4 weeks [4]. A comparison of diagnoses made by primary care physicians and those made by laryngologists and speech-language pathologists with experience in interpreting stroboscopy at multidisciplinary voice clinics indicated that the primary care physicians' diagnoses of dysphonia were different in 45%-70% of cases [4]. However, the laryngoscopic examination is an invasive procedure. To achieve an accurate diagnosis, it must be performed by an experienced laryngologist. The examination equipment is expensive and not generally available in primary care units. In places without sufficient medical resources, delayed diagnoses and treatments are common [5]. Therefore, a noninvasive diagnostic tool is needed to resolve this problem. Although this tool cannot replace the laryngoscopic examination by an experienced physician, it is worthwhile to develop because a noninvasive tool to screen significant clinical conditions could encourage patients to visit a voice clinic for further evaluation.

Several recent studies have attempted to distinguish normal and abnormal voices by using various machine learning–based classifiers that have the potential for detecting pathological voices [5-9]. To date, the highest accuracy of pathological voice detection achieved by using a deep neural network has been 99.32% [5]. However, the differential diagnosis of various types of pathological voices has not been widely reported. The vibration patterns of vocal fold observed by high-speed video for common vocal fold diseases, including vocal atrophy, unilateral vocal paralysis, and organic vocal fold lesions, are completely different [10]. We hypothesized that different vibration patterns could result in different voice features. This study sought to detect various vocal fold diseases through pathological voice recognition using a deep learning approach.

Methods

Sample Collection

This study was performed following the principles expressed in the Declaration of Helsinki, and approved by the Institutional Ethics and Research Committee of Cheng Hsin General Hospital and Fu Jen Catholic University. Voice samples were obtained from the Voice Center of Chen Hsin General Hospital and the Department of Otorhinolaryngology-Head and Neck Surgery of Fu Jen Catholic University Hospital. These samples included 189 normal voice samples and 552 samples of voice disorders, including vocal atrophy (n=224), unilateral vocal paralysis (n=50), organic vocal fold lesions (n=248), and adductor spasmodic dysphonia (n=30). Voice samples of a sustained vowel sound /a:/ followed by continuous speech of a Mandarin passage [11] (Multimedia Appendix 1) were recorded at a comfortable loudness level with a microphone-to-mouth distance of approximately 15-20 cm using a high-quality microphone with a digital amplifier and a 40- to 45-dB background noise level. The sampling rate was 44,100 Hz with 16-bit resolution, and data were saved in an uncompressed .wav format.

Comparison and Evaluation

We first divided the 741 samples into 2 sets: 593 samples for the training set and 148 samples for the testing set. Using computer-based randomization, we selected 152 of the 189 normal voice samples, 40 of the 50 unilateral vocal paralysis samples, 24 of the 30 adductor spasmodic dysphonia samples, 179 of the 224 vocal atrophy samples, and 198 of the 248 organic vocal fold lesion samples for the training set (Table 1).

Table 1. Details of the voice samples used for experiments (N=741).

Sample	Training set (n=593)	Test set (n=148)
Normal	152	37
Unilateral vocal paralysis	40	10
Adductor spasmodic dysphonia	24	6
Vocal atrophy	179	45
Organic vocal fold lesions	198	50

To manage the limited size of the training set, we used a mix-up approach for data augmentation [12]. The mix-up approach has been applied for audio scene classification using convoluted neural networks (CNNs) to reduce overfitting and obtain higher prediction accuracy [13]. We randomly selected 2 voice files and mixed them into 1 voice file with randomly selected weights to construct the virtual training examples. Next, we randomly cropped each of these voice files to achieve 10 voice files with a length of 11.88 seconds (plateau point of the training length within the graphics processing unit memory limitations of our hardware, according to our preliminary tests). Additionally, we used oversampling to adjust the class distribution of the data [14].

A 2D graph is ideal for extracting features when using CNNs. Therefore, we performed Mel frequency cepstral coefficients (MFCCs) for the processed voice file to obtain a spectrogram. Feature extraction from MFCCs was performed using pre-emphasis, windowing, fast Fourier transform, Mel filtering, nonlinear transformation, and discrete cosine transform [15]. The first feature consisted of 40-dimension MFCCs [16,17]. Next, for the second and third features, we calculated the MFCC trajectories over time (delta MFCCs) and the second-order delta of MFCCs. Therefore, there were 3 channels of input features that could be considered a color image (ie, red–green–blue in the computer vision field).

CNNs have distinct feature representation–related characteristics, among which the lower layers provide general feature-extraction capabilities and the higher layers include information that is increasingly more specific to the original classification task [18]. This allows verbatim reuse of the generalized feature-extraction and representation of the lower CNN layers; the higher layers are fine-tuned toward secondary problem domains with characteristics related to the original. Therefore, instead of designing a new CNN with random parameter initialization, it is more suitable to adopt a pretrained CNN and fine-tune its parameterization toward specific classification domains. Spectrograms were quite different from normal images at first glance. However, the low-level features, including edges, corners, and shapes, were common in the normal images and spectrograms [19]. In a previous study, a spectrogram-based crowd sounds analysis using pretrained CNN models from the ImageNet data set showed great accuracy when distinguishing crowd emotions [19]. Another study also proved that pretrained CNN models yielded better performance than nontrained CNN models for classifying normal or pathological cases [18]. We used different CNN architectures, such as EfficientNet-B0 to B6 [20], SENet154 [21], Se_resnext101_32x4d [21], and se_resnet152 [21] models, from the ImageNet data set that have been pretrained for transfer learning. We classified pathological conditions into 2 (normal voice; adductor spasmodic dysphonia plus organic vocal fold lesions plus unilateral vocal paralysis plus vocal atrophy), 3 (normal voice; adductor spasmodic dysphonia; organic vocal fold lesions plus unilateral vocal paralysis plus vocal atrophy), 4 (normal voice; adductor spasmodic dysphonia; organic vocal fold lesions; unilateral vocal paralysis plus vocal atrophy), or 5 (normal voice; adductor spasmodic dysphonia; organic vocal fold lesions; unilateral vocal paralysis; vocal atrophy) different conditions and trained the CNN. For the final prediction of an input instance, we used the maximum probability to obtain the label.

In terms of hyperparameter settings for fine-tuning among the training set, 474 of 593 samples (79.9%) were used for initial training and 119 of 593 samples (20.1%) were used for validation. We added the dropout function and different data augmentation methods to prevent the model from overfitting in our data set [22,23]. The dropout rate was set at 0.25-0.5 for regularization. Then, we trained the model using minibatches of 32 that were selected based on memory consumption [24]. The learning rate was tuned based on cosine annealing and a 1–cycle policy strategy [25,26]. By using the cosine annealing schedule, the model repeatedly fitted the gradient to the local minimum. The network was trained end-to-end using the Adam optimization algorithm, and it optimized the cross-entropy as a loss function [27]. For different classification problems in the model head, we applied a SoftMax layer as an output layer for multiclass classification or a sigmoid layer for binary classification. Finally, we assembled the model by average output probability to receive more robust results to minimize the bias of prediction error to improve the prediction accuracy of the CNN models [25]. The machine learning process was performed using Python 3.8 (Python Software Foundation) and PyTorch 1.7.1 for Ubuntu 18.04 (Facebook's AI Research lab [FAIR]). Furthermore, we invited 2 laryngologists and 2 general ear, nose, and throat (ENT) physicians who could speak Mandarin to categorize the voice samples of the testing sets into 5 classifications. We compared their classifications with those of our model.

Statistical Analysis

The effectiveness of our model was evaluated by several metrics, including accuracy, sensitivity, specificity, F1 score, receiver-operating characteristic (ROC) curve, and area under the curve (AUC). All metrics were calculated using Python.

Results

Voice samples in this study were composed of a sustained vowel sound and a continuous essay speech. We first compared the results by training different parts: the vowel sound alone, the essay alone, and the whole voice sample (ie, the vowel sound and essay). Because the vowel sound and essay group (F1 score=0.65) achieved better F1 scores than the vowel sound group (F1 score=0.54) and the essay group (F1 score =0.57), we applied whole voice samples during subsequent machine learning. Figure 1 shows the changes in the loss function value over the training and validation sets, which demonstrated that our model could converge after running the optimization for a number of epochs.

Figure 1. Illustration of the changes of the loss function value over the training and validation sets.

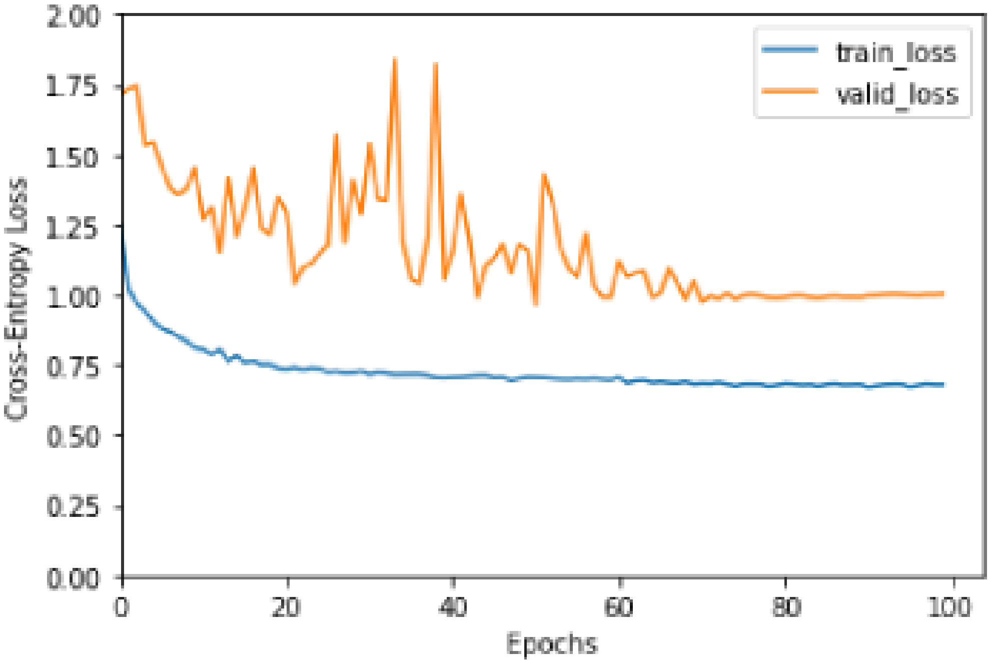


Table 2 presents the training results for the different classification methods, including 2 (normal voice; adductor spasmodic dysphonia plus organic vocal fold lesions plus unilateral vocal paralysis plus vocal atrophy), 3 (normal voice; adductor spasmodic dysphonia; organic vocal fold lesions plus unilateral vocal paralysis plus vocal atrophy), 4 (normal voice; adductor spasmodic dysphonia; organic vocal fold lesions; unilateral vocal paralysis; vocal atrophy), or 5 (normal voice; adductor spasmodic dysphonia; organic vocal fold lesions; unilateral vocal paralysis; vocal atrophy) different conditions trained by the CNN. The 2-classification condition could equally distinguish pathological voices from normal voices. In our model, the accuracy of pathological voice detection reached 95.3%; the sensitivity was 99%, specificity was 84%, and AUC was 0.98. Using the 3-classification condition, we aimed to identify adductor spasmodic dysphonia patients from those with other vocal fold pathologies. The accuracy was 91.2%, sensitivity was 82%, specificity was 93%, and AUC was 0.91. Using the 4-classification condition, vocal atrophy and unilateral vocal paralysis could be clinically grouped as "glottic insufficiency." For this condition, the accuracy was 71.0%, sensitivity was 75%, specificity was 89%, and AUC was 0.88. Using the 5-classification condition, the accuracy was 66.9%, sensitivity was 66%, specificity was 91%, and AUC was 0.85. Figure 2 shows the confusion matrix of these results. Figure 3 shows the ROC curves of these results.

Table 2. Performance of the artificial intelligence model for classifying voice disorders under different classification conditions.

Class	Sensitivity	Specificity	Accuracy, %	F1 score	Average area under the curve values
2	0.99	0.84	95.3	0.97	0.98
3	0.82	0.93	91.2	0.80	0.96
4	0.75	0.89	71.0	0.75	0.88
5	0.66	0.91	66.9	0.66	0.85

Figure 2. Confusion matrix of 2, 3, 4, and 5 classifications. AN = pathological voice; NC = normal voice; SD = adductor spasmodic dysphonia; PAATOL = unilateral vocal paralysis/vocal atrophy/organic vocal fold lesions; OL = organic vocal fold lesions; PAAT = unilateral vocal paralysis/vocal atrophy; PA = unilateral vocal paralysis; AT = vocal atrophy.

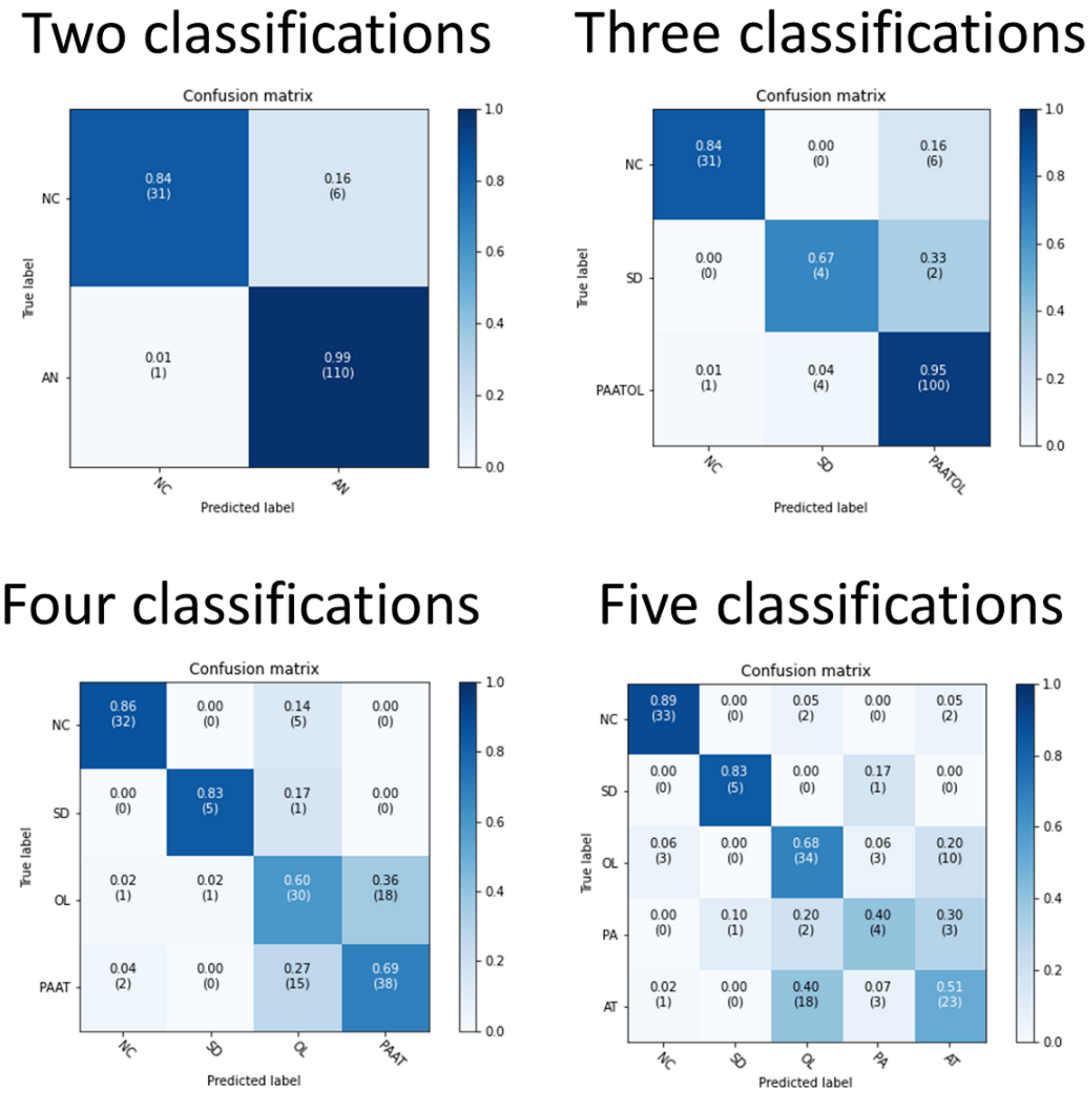
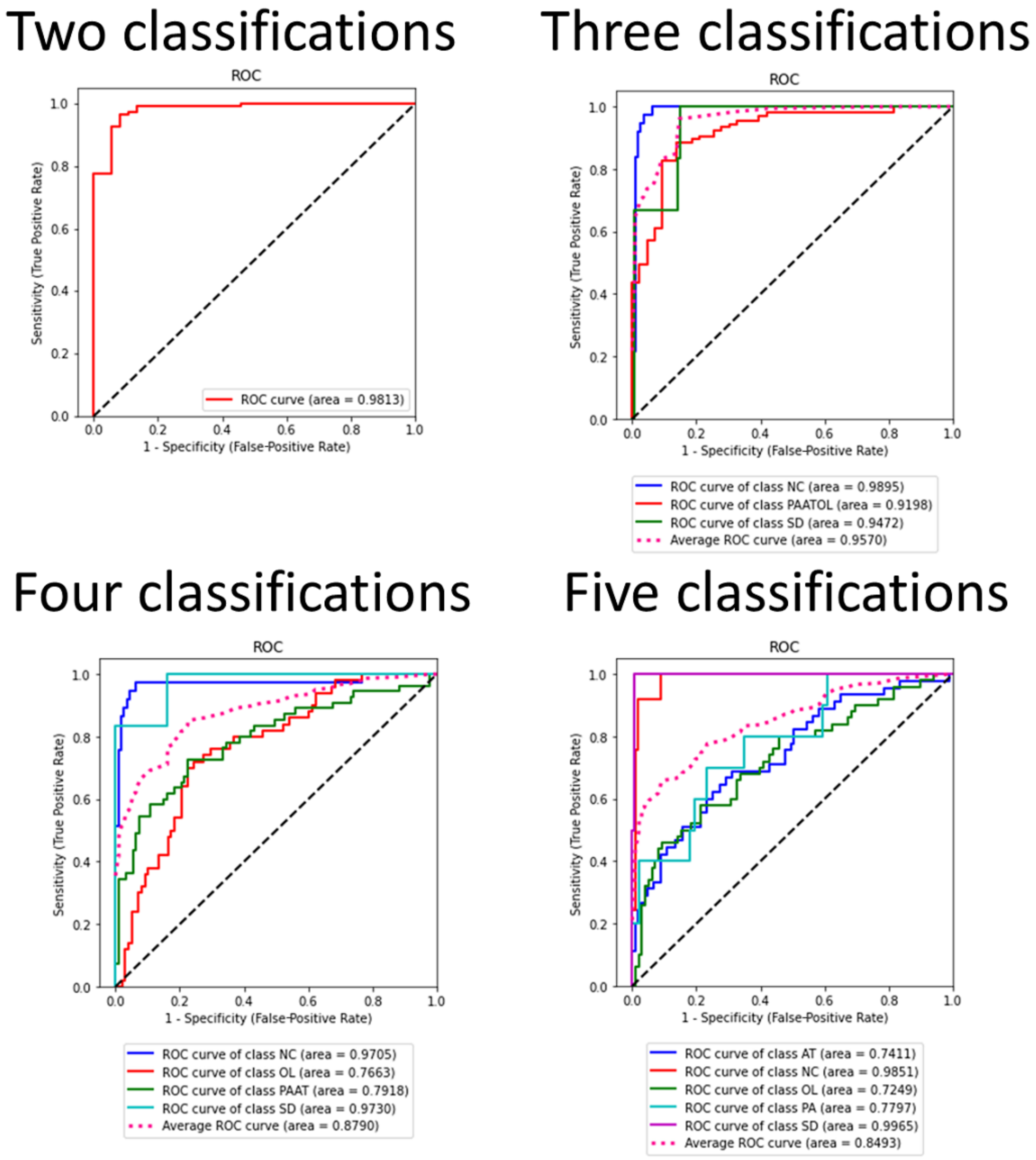


Figure 3. Receiver operating characteristic curves of 2, 3, 4, and 5 classifications. NC = normal voice; SD = adductor spasmodic dysphonia; PAATOL = unilateral vocal paralysis/vocal atrophy/organic vocal fold lesions; OL = organic vocal fold lesions; PAAT = unilateral vocal paralysis/vocal atrophy; PA = unilateral vocal paralysis; AT = vocal atrophy.



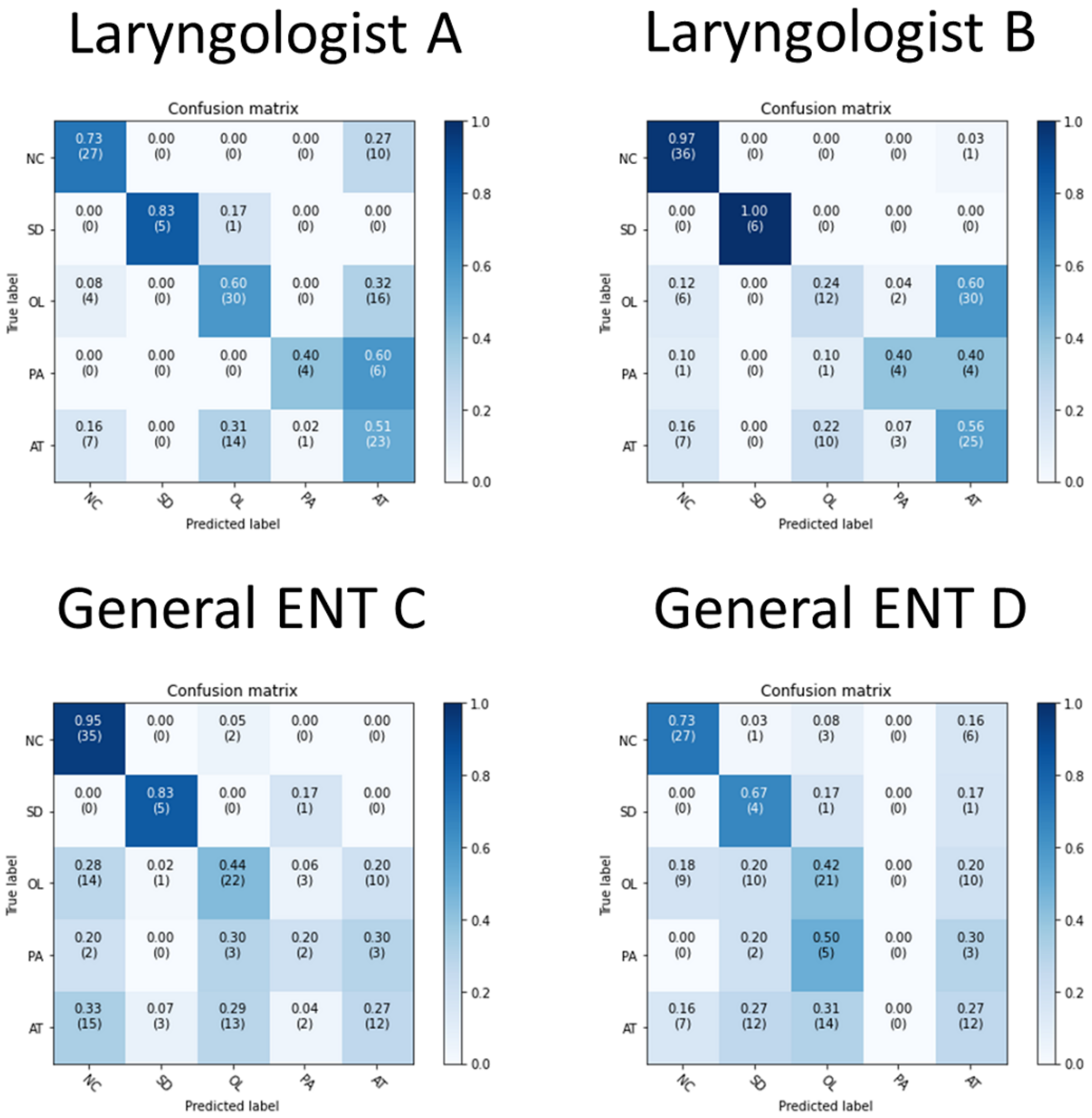
Furthermore, we invited four ENT specialists to identify vocal fold pathology by voice using these 5 classifications. The results are shown in Table 3 and Figure 4. The accuracy rates were 60.1% and 56.1% for the 2 laryngologists and 51.4% and 43.2% for the 2 general ENT specialists.

Table 3. Comparison of the performance for a 5-classification condition by our artificial intelligence model and 4 human experts.

Test participants	Sensitivity	Specificity	Accuracy, %
Deep learning model	0.66	0.91	66.9
Laryngologist A (11 years of experience)	0.61	0.89	60.1
Laryngologist B (10 years of experience)	0.63	0.88	56.1
General ENT[a] C (8 years of experience)	0.54	0.88	51.4
General ENT D (14 years of experience)	0.42	0.85	43.2

[a] ENT: ear, nose, and throat.

Figure 4. Confusion matrix of 5 classifications in human specialists. NC = normal voice; SD = adductor spasmodic dysphonia; OL = organic vocal fold lesions; PA = unilateral vocal paralysis; AT = vocal atrophy.



Discussion

Principal Findings

During this study, we built a CNN model that could distinguish, with high specificity (91%), different pathological voices attributable to common vocal diseases based on voice alone. To the best of our knowledge, no previous study has used artificial intelligence to distinguish different types of pathological voices speaking Mandarin. Using our model, we obtained better results by training the CNN with a whole voice sample than by training it with the vowel sound only or with the essay speech only.

Our model could distinguish normal voice and adductor spasmodic dysphonia with great performance for the 5-classification condition (AUC values: 0.985 and 0.997, respectively). The overall accuracy of our model was also better than that of all ENT specialists participating in the study. This was compatible with our clinical observation that the first impression of the pathological voice is usually over-ruled by the laryngoscopic examination. Additionally, laryngologists demonstrated higher accuracy when diagnosing voice disorders than general ENT specialists. This may imply that it would be possible to improve the accuracy of human physicians in terms of their impressions of pathological voices by increasing clinical experience. After comparing the accuracy of each classification, we found that artificial intelligence was markedly better than laryngologists when identifying organic vocal fold lesions (artificial intelligence, 68%; laryngologist A, 60%; laryngologist B, 24%). However, laryngologists were slightly better at vocal atrophy identification (artificial intelligence, 51%; laryngologist A, 51%; laryngologist B, 56%).

Organic vocal fold lesions, unilateral vocal paralysis, and vocal atrophy could result in a closure gap during phonation, inducing a weak and breathy sound [28-30], and vocal fold tension imbalance, inducing diplophonia (when a voice is perceived as being produced with 2 concurrent pitches) [31]. Specifically, in the case of organic vocal fold lesions during vibration, the lesion divided the fold into 2 oscillators. However, in the case of unilateral vocal paralysis, vibrating frequencies were different between the normal vocal fold and paralysis vocal fold. Vocal atrophy will show a breakdown of vibration with a visible repetition in the loss of normal vibration every few glottal cycles [10]. However, the difference in the vibration pattern could only be observed by high-speed video and multislice digital videokymography [10], and the resulting pathological voice is difficult for humans to identify. We speculated that our model may identify related features through deep learning to achieve better outcomes.

Laryngologists could distinguish aged and young patients, and they could validate their judgment during the test based on their knowledge. Vocal atrophy is the most common vocal fold pathology in older patients [3]. Therefore, laryngologists may classify the pathological voice as vocal atrophy if they judged that the voice was that of an aged person.

Regarding misclassification, we have found that our model could successfully identify normal voice and spasmodic dysphonia. However, it was relatively difficult to differentiate organic vocal lesions, unilateral vocal paralysis, and vocal atrophy from each other. Although the vibration patterns were different for these 3 diseases, the different severity levels of disease could result in different degrees of hoarseness. For example, with tiny vocal nodules compared with huge vocal polyps, unilateral vocal paralysis with fair compensation compared with unilateral vocal paralysis with a huge closure gap, and vocal atrophy with a mild anterior closure gap compared with vocal atrophy with a huge closure gap, there could be different degrees of hoarseness in the same group. We assumed that the less severe cases in each group may not show the typical pathological vibration pattern. Further studies are needed to validate our hypothesis.

Four human specialists required 40-80 minutes to identify 148 voice samples of the test set; however, our model only required 30 seconds to perform the same task. The processing time of our model is quite promising in terms of the development of future screening tools.

Comparison With Prior Work

Most previous studies have used sustained vowel sounds for pathological voice detection [5,7,8]. However, other studies have used continuous speech samples for analyses [6]. CNNs extract features automatically from the spectrogram of voice recordings for dysphonia diagnosis, and a larger amount of training data yields better results [32]. Therefore, the CNN used here may have extracted more features from these entire voice samples, thereby achieving better training results with our model.

In this study, we used our voice database for the deep learning approach. The most widely used voice disorder database is the Massachusetts Eye & Ear Infirmary (MEEI) Voice Disorders Database (commercially available from KayPENTAX Inc.). The MEEI voice samples (53 normal and 662 pathological voices) are composed of the vowel /ah/ (53 normal and 657 pathological voices) and the utterance of a sentence ("When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow") [33]. However, the voice recordings in the MEEI database were recorded at various sampling rates (10, 25, and 50 kHz), and normal and pathological voice recordings were recorded in 2 different environments [32]. Therefore, it was not clear whether artificial intelligence was classifying voice features or environments when trained using the MEEI samples.

The other widely used voice disorder database is the Saarbruecken Voice Database, which contains voice recordings from more than 2000 individuals. Each participant file contains recordings of sustained vowel sounds of /a/, /i/, and /u/ in low, neutral, high, and low–high–low pitches, as well as a continuous speech sentence ("Guten Morgen, wie geht as Ihnen?"). All these samples were recorded at 50-kHz sampling rates and 16-bit resolution [32]. The Saarbruecken Voice Database is considered to be superior to the MEEI database because it uses the same recording environment and the same sampling rates. However, it contains 71 different dysphonia pathologies and many patients recorded in this database had multiple disorders. Therefore, it is difficult to achieve denotation before machine learning.

Our database has several advantages. First, all voice data were from patients visiting our clinics who had detailed chart documents that were carefully reviewed by 2 experienced laryngologists (H-CH and S-YC). Therefore, the quality of the primary data was better than that of the primary data of other studies during which voice data were retrieved from a public database. Second, all voice data were recorded using 44.1-kHz sampling rates and 16-bit resolution, which comprise the standard audio CD format. This widely used format could increase the usability of this data set. Third, we focused on 4 vocal fold diseases that were chosen by experienced laryngologists based on the cause of hoarseness, clinical significance, and prevalence of the disease.

In terms of the cause of hoarseness, adductor spasmodic dysphonia is a focal laryngeal dystonia characterized by irregular and uncontrolled voice breaks that interrupt normal speech [34]. However, organic vocal fold lesions, unilateral vocal paralysis, and vocal atrophy could induce a breathy sound with a different diplophonia pattern [10]. The voice pattern of adductor spasmodic dysphonia is quite different from that of the other pathologies included in this study, and classic cases of adductor spasmodic dysphonia could be diagnosed based on voice alone by experienced laryngologists. The accuracy rates of adductor spasmodic dysphonia among laryngologists using the 5-classification condition were 100% and 83% during this study. Therefore, we anticipated that the sensitivity and specificity for diagnosing adductor spasmodic dysphonia could be higher than those of other categories. However, during the first attempt at CNN training, the accuracy of adductor spasmodic dysphonia identification was poor (data not shown). When we attempted to splice the original voice file into 1-second clips while training the model, we found that the voice break in adductor spasmodic dysphonia did not always emerge within every 1-second period.

After prolonging the duration of the voice clips in the training model, the results improved substantially. This also emphasized that the domain knowledge could significantly influence the training results by tuning the training model according to real clinical conditions.

According to a meta-analysis, in terms of clinical significance, patients with neurologic voice disorders have more challenges than patients with inflammatory or traumatic laryngeal diseases [1]. Specifically, adductor spasmodic dysphonia showed the worst Voice Handicap Index (VHI) score, followed by unilateral vocal paralysis [1]. This result was compatible with our clinical observation that adductor spasmodic dysphonia could markedly interfere with communication and socialization during the daily lives of patients. Although adductor spasmodic dysphonia is a rare disease with a prevalence of 14 out of 100,000 [35], it is worthwhile to offer a model for rapid screening because the symptoms can be treated easily and effectively by regular intralaryngeal botulinum toxin type A injections or surgery [35].

According to VHI scores, unilateral vocal paralysis could also induce a severe voice handicap [1]. The most common cause of unilateral vocal paralysis is an idiopathic or postviral infection, which accounts for 67% of cases [36]. However, 6% of patients have underlying malignancies that invade the recurrent laryngeal nerve or vagus nerve [36]. Computed tomography of the skull base, neck, and chest is often recommended during the search for a potential cause of the voice disorder [37]. Thyroid disease, including benign nodules, thyroid malignancy, thyroiditis, hyperthyroidism, and hypothyroidism, may also result in vocal fold paresis [38]. Heman-Ackah et al [38] reported that 47.4% of patients with unilateral vocal paralysis are diagnosed with concurrent thyroid disease. Therefore, it is important to determine an early diagnosis of unilateral vocal paralysis to investigate the existence of underlying disease.

Organic vocal fold lesions comprise benign lesions, such as nodules, polyps, cysts, polypoid vocal folds, precancerous leukoplakia, and malignant lesions [39,40]. The cause of hoarseness with benign and malignant vocal fold lesions involves changes in the laryngeal mucosa and mass effects [10,40]. To date, it has been difficult to differentiate organic vocal fold lesions further by voice alone because they involve various pathologies. However, it is worthwhile to inform patients about the possibility of organic vocal fold lesions and to advise them to undergo further investigations. Early stage malignant lesions and benign lesions could be treated with office-based surgery, which is safer and relatively inexpensive compared with surgery in the operating room [29,41,42]

The most common cause of vocal atrophy is aging. Aging may result in atrophic musculature and a thinner lamina propria of the vocal fold [43]. However, vocal atrophy can also occur in a relatively young population [28]. It may result from a congenital anomaly or prolonged laryngopharyngeal reflux [44]. The symptoms associated with vocal atrophy are relatively subtle compared with those of other vocal fold diseases [45]. The concern about significant underlying diseases is also reduced with vocal atrophy. However, vocal atrophy is the most common vocal fold pathology among patients older than 65 years [3]. With the aging of the population, vocal atrophy may become a significant geriatric issue in the future.

Limitations

Our study had some drawbacks. First, all the voice files were recorded in the studio with a silent environment, with sensitive audio-recording technology, and using a certain format. Further studies are needed to validate this approach in different recording environments. Second, all the voice data in this study are from Mandarin speakers. Further studies are needed to compare the results of speakers of different languages. Third, the voice sample numbers of each class were unequal because of the different disease prevalence. We have applied several data augmentation methods to ameliorate the influence of these unequal data. Fraile et al [46] showed that laryngeal pathology detection using voice records based on MFCC and prior differentiation by sex can significantly improve the performance. Fang et al [47] also showed that a deep neural network combining supervectors with medical records could improve pathological voice classifications. Therefore, in the future, we will combine demographic data with voice records to improve our model.

Conclusions

We demonstrated that voice alone could be used for common vocal fold disease recognition using a deep learning application after training with our Mandarin pathological voice database. Specifically, adductor spasmodic dysphonia, organic vocal fold lesions, unilateral vocal paralysis, and vocal atrophy could be recognized, which could increase the potential of this approach to be more beneficial than simply distinguishing a pathological voice from a normal voice. This approach shows clinical potential for use during general screening of different vocal fold diseases based on voice and could be included in quick evaluations during general health examinations. It could also be used for telemedicine in remote regions that lack laryngoscopy services in primary care units. Overall, it could support physicians during prescreening of cases by allowing for invasive examinations to be performed only for cases involving problems with automatic recognition or listening and for professional analyses of other clinical examination results that reveal doubts about the presence of pathologies.

Acknowledgments

The authors acknowledge financial support from the Ministry of Science and Technology (MOST 109-2926-I-010-501, MOST 107-2314-B-010-015-MY3, MOST 109-2926-I-010-502, MOST 109-2321-B-010-005, MOST 108-2923-B-010-002-MY3, MOST 109-2823-8-010-003-CV, MOST 109-2622-B-010-006, and MOST 109-2321-B-010-006) and Fu Jen Catholic University Hospital (PL-201908012-V and PL-202008005-V). This work was particularly supported by “Development and Construction Plan” of the

School of Medicine, National Yang-Ming University, which is now known as National Yang Ming Chiao Tung University (107F-M01-0504), and Aiming for the Top University Plan, a grant from Ministry of Education. Funders did not have any role in the study design, data collection and analysis, the decision to publish, or manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mandarin passage.
[DOCX File , 12 KB - jmir_v23i6e25247_app1.docx]

References

1. Cohen SM, Dupont WD, Courey MS. Quality-of-life impact of non-neoplastic voice disorders: a meta-analysis. Ann Otol Rhinol Laryngol 2006 Feb;115(2):128-134. [doi: 10.1177/000348940611500209] [Medline: 16514796]
2. Cohen SM, Kim J, Roy N, Asche C, Courey M. Prevalence and causes of dysphonia in a large treatment-seeking population. Laryngoscope 2012 Feb;122(2):343-348. [doi: 10.1002/lary.22426] [Medline: 22271658]
3. Davids T, Klein AM, Johns MM. Current dysphonia trends in patients over the age of 65: is vocal atrophy becoming more prevalent? Laryngoscope 2012 Feb;122(2):332-335. [doi: 10.1002/lary.22397] [Medline: 22252988]
4. Stachler RJ, Francis DO, Schwartz SR, Damask CC, Digoy GP, Krouse HJ, et al. Clinical Practice Guideline: Hoarseness (Dysphonia) (Update). Otolaryngol Head Neck Surg 2018 Mar;158(1_suppl):S1-S42. [doi: 10.1177/0194599817751030] [Medline: 29494321]
5. Fang SH, Tsao Y, Hsiao MJ, Chen JY, Lai YH, Lin FC, et al. Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach. J Voice 2019 Sep;33(5):634-641. [doi: 10.1016/j.jvoice.2018.02.003] [Medline: 29567049]
6. Umapathy K, Krishnan S, Parsa V, Jamieson DG. Discrimination of pathological voices using a time-frequency approach. IEEE Trans Biomed Eng 2005 Mar;52(3):421-430. [doi: 10.1109/TBME.2004.842962] [Medline: 15759572]
7. Godino-Llorente JI, Gómez-Vilda P, Blanco-Velasco M. Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. IEEE Trans Biomed Eng 2006 Oct;53(10):1943-1953. [doi: 10.1109/TBME.2006.871883] [Medline: 17019858]
8. Costa S, Neto B, Fechine J. Pathological voice discrimination using cepstral analysis, vector quantization and hidden Markov models. 2008 Presented at: 8th IEEE International Conference on BioInformatics and BioEngineering; October 8-10, 2008; Athens, Greece. [doi: 10.1109/bibe.2008.4696783]
9. Salhi L, Mourad T, Cherif A. Voice disorders identification using multilayer neural network. Int Arab J Inf Technol 2010 Apr;2(7):177-185 [FREE Full text]
10. Woo P. Vibratory Characteristics of Diplophonia Studied by High Speed Video and Vibrogram Analysis. J Voice 2019 Jan;33(1):7-15. [doi: 10.1016/j.jvoice.2017.08.013] [Medline: 30389188]
11. Chen SH. The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. J Acoust Soc Am 2005 May;117(5):3225-3230. [doi: 10.1121/1.1872312] [Medline: 15957789]
12. Zhang H, Cisse M, Dauphin Y, Lopez-Paz D. Mixup: beyond empirical risk minimization. arXiv 2018 Apr [FREE Full text]
13. Xu K, Feng D, Mi H, Zhu B, Wang D, Zhang L, et al. Mixup-based acoustic scene classification using multi-channel convolutional neural network. Switzerland: Springer Nature; 2018 Presented at: 19th Pacific-Rim Conference on Multimedia; 2018 Sep 21-22; Hefei, China.
14. Branco P, Torgo L, Ribeiro R. A survey of predictive modelling under imbalanced distributions. arXiv 2015 May [FREE Full text]
15. Logan B. Mel Frequency cepstral coefficients for music modeling. 2000 Presented at: 1st International Symposium on Music Information Retrieval; October 23-25, 2000; Plymouth, MA URL: https://ismir2000.ismir.net/papers/logan_paper.pdf
16. Ravi Kumar KM, Ganesan S. Comparison of multidimensional MFCC feature vectors for objective assessment of stuttered disfluencies. Int J Adv Netw Appl 2011;2(5):854-860 [FREE Full text]
17. Peng N, Chen A, Zhou G, Chen W, Zhang W, Liu J, et al. Environment Sound Classification Based on Visual Multi-Feature Fusion and GRU-AWS. IEEE Access 2020;8:191100-191114. [doi: 10.1109/access.2020.3032226]
18. Hussain M, Bird J, Faria D. A study on CNN transfer learning for image classification. 2018 Presented at: 18th Annual UK Workshop on Computational Intelligence; September 5-7, 2018; Nottingham, UK. [doi: 10.1007/978-3-319-97982-3_16]
19. Franzoni V, Biondi G, Milani A. Emotional sounds of crowds: spectrogram-based analysis using deep learning. Multimed Tools Appl 2020 Aug 17;79:36063-36075. [doi: 10.1007/s11042-020-09428-x]
20. Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv 2020 Sep [FREE Full text]
21. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2018 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: 10.1109/cvpr.2018.00745]

22. Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. arXiv 2017 Dec [FREE Full text]
23. Wu H, Gu X. Towards dropout training for convolutional neural networks. Neural Netw 2015 Nov;71:1-10. [doi: 10.1016/j.neunet.2015.07.007] [Medline: 26277608]
24. Lym S, Behroozi A, Wen W, Li G, Kwon Y, Erez M. Mini-batch serialization: CNN training with inter-layer data reuse. arXiv 2018 May [FREE Full text]
25. Huang G, Li Y, Pleiss G, Liu Z, Hopcroft J, Weinberger K. Snapshot ensembles: train 1, get M for free. arXiv 2017 Apr 01 [FREE Full text]
26. Smith L. A disciplined approach to neural network hyper-parameters: Part 1 - learning rate, batch size, momentum, and weight decay. arXiv 2018 Apr 24 [FREE Full text]
27. Kingma D, Ba J. Adam: A method for stochastic optimization. arXiv 2017 Jan 30 [FREE Full text]
28. Hu HC, Hung YT, Lin SY, Tung TH, Chang SY. Office-Based Autologous Fat Injection Laryngoplasty for Glottic Insufficiency in Patients Under 50 Years Old. J Voice 2019 Sep;33(5):747-750. [doi: 10.1016/j.jvoice.2018.03.012] [Medline: 29678440]
29. Hu HC, Hung YT, Lin SY, Chang SY. Office-based Autologous Fat Injection Laryngoplasty for Vocal Process Granuloma. J Voice 2016 Nov;30(6):758.e7-758.e11. [doi: 10.1016/j.jvoice.2015.09.003] [Medline: 26454769]
30. Johns MM, Arviso LC, Ramadan F. Challenges and opportunities in the management of the aging voice. Otolaryngol Head Neck Surg 2011 Jul;145(1):1-6. [doi: 10.1177/0194599811404640] [Medline: 21521897]
31. Hong KH, Kim HK. Diplophonia in unilateral vocal fold paralysis and intracordal cyst. Otolaryngol Head Neck Surg 1999 Dec;121(6):815-819. [doi: 10.1053/hn.1999.v121.a94214] [Medline: 10580244]
32. Wu H, Soraghan J, Lowit A, Di Caterina G. Convolutional Neural Networks for Pathological Voice Detection. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:1-4. [doi: 10.1109/EMBC.2018.8513222] [Medline: 30440307]
33. Daoudi K, Bertrac B. On classification between normal and pathological voices using the MEEI-KayPENTAX database: issues and consequences. 2014 Presented at: 15th Annual Conference of the International Speech Communication Association; September 14-18, 2014; Singapore URL: https://www.isca-speech.org/archive/archive_papers/interspeech_2014/i14_0198.pdf
34. Hintze JM, Ludlow CL, Bansberg SF, Adler CH, Lott DG. Spasmodic Dysphonia: A Review. Part 1: Pathogenic Factors. Otolaryngol Head Neck Surg 2017 Oct;157(4):551-557. [doi: 10.1177/0194599817728521] [Medline: 28850801]
35. Whurr R, Lorch M. Review of differential diagnosis and management of spasmodic dysphonia. Curr Opin Otolaryngol Head Neck Surg 2016 Jun;24(3):203-207. [doi: 10.1097/MOO.0000000000000253] [Medline: 26900821]
36. Koufman JA, Postma GN, Cummins MM, Blalock PD. Vocal fold paresis. Otolaryngol Head Neck Surg 2000 Apr;122(4):537-541. [doi: 10.1067/mhn.2000.102574] [Medline: 10740174]
37. Ivey CM. Vocal Fold Paresis. Otolaryngol Clin North Am 2019 Aug;52(4):637-648. [doi: 10.1016/j.otc.2019.03.008] [Medline: 31088695]
38. Heman-Ackah YD, Joglekar SS, Caroline M, Becker C, Kim E, Gupta R, et al. The prevalence of undiagnosed thyroid disease in patients with symptomatic vocal fold paresis. J Voice 2011 Jul;25(4):496-500. [doi: 10.1016/j.jvoice.2010.03.008] [Medline: 20728306]
39. Naunheim MR, Carroll TL. Benign vocal fold lesions: update on nomenclature, cause, diagnosis, and treatment. Curr Opin Otolaryngol Head Neck Surg 2017 Dec;25(6):453-458. [doi: 10.1097/MOO.0000000000000408] [Medline: 29099730]
40. Rzepakowska A, Sielska-Badurek E, Cruz R, Sobol M, Osuch-Wójcikiewicz E, Niemczyk K. Narrow band imaging versus laryngovideostroboscopy in precancerous and malignant vocal fold lesions. Head Neck 2018 May;40(5):927-936. [doi: 10.1002/hed.25047] [Medline: 29322573]
41. Hu HC, Lin SY, Hung YT, Chang SY. Feasibility and Associated Limitations of Office-Based Laryngeal Surgery Using Carbon Dioxide Lasers. JAMA Otolaryngol Head Neck Surg 2017 May 01;143(5):485-491 [FREE Full text] [doi: 10.1001/jamaoto.2016.4129] [Medline: 28208177]
42. Hsu YB, Lan MC, Chang SY. Percutaneous corticosteroid injection for vocal fold polyp. Arch Otolaryngol Head Neck Surg 2009 Aug;135(8):776-780. [doi: 10.1001/archoto.2009.86] [Medline: 19687397]
43. Kendall K. Presbyphonia: a review. Curr Opin Otolaryngol Head Neck Surg 2007 Jun;15(3):137-140. [doi: 10.1097/MOO.0b013e328166794f] [Medline: 17483679]
44. Habesoglu TE, Habesoglu M, Sürmeli M, Deveci I, Toros SZ, Gunes P, et al. Histological changes of rat soft palate with exposure to experimental laryngopharyngeal reflux. Auris Nasus Larynx 2010 Dec;37(6):730-736. [doi: 10.1016/j.anl.2010.03.009] [Medline: 20413235]
45. Gartner-Schmidt J, Rosen C. Treatment success for age-related vocal fold atrophy. Laryngoscope 2011 Mar;121(3):585-589. [doi: 10.1002/lary.21122] [Medline: 21344441]
46. Fraile R, Sáenz-Lechón N, Godino-Llorente JI, Osma-Ruiz V, Fredouille C. Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex. Folia Phoniatr Logop 2009;61(3):146-152 [FREE Full text] [doi: 10.1159/000219950] [Medline: 19571549]
47. Fang SH, Wang CT, Chen JY, Tsao Y, Lin FC. Combining acoustic signals and medical records to improve pathological voice classification. APSIPA Transactions on Signal and Information Processing 2019;8(e14):1-11. [doi: 10.1017/atsip.2019.7]

Abbreviations

AUC: area under the curve
CNNs: convoluted neural networks
ENT: ear, nose, and throat
MEEI: Massachusetts Eye & Ear Infirmary
MFCCs: Mel frequency cepstral coefficients
ROC: receiver-operating characteristic
VHI: Voice Handicap Index

Edited by R Kukafka; submitted 25.10.20; peer-reviewed by YB Hsu, V Franzoni; comments to author 09.12.20; revised version received 06.04.21; accepted 25.04.21; published 08.06.21.

Please cite as:
Hu HC, Chang SY, Wang CH, Li KJ, Cho HY, Chen YT, Lu CJ, Tsai TP, Lee OKS
Deep Learning Application for Vocal Fold Disease Prediction Through Voice Recognition: Preliminary Development Study
J Med Internet Res 2021;23(6):e25247
URL: https://www.jmir.org/2021/6/e25247
doi:10.2196/25247
PMID:34100770

Original Paper

The Accuracy of Wrist Skin Temperature in Detecting Ovulation Compared to Basal Body Temperature: Prospective Comparative Diagnostic Accuracy Study

Tracy Y Zhu[1], DPhil; Martina Rothenbühler[2], DPhil; Györgyi Hamvas[1], MSc; Anja Hofmann[1], MA; JoEllen Welter[1], MPH; Maike Kahr[1], MD; Nina Kimmich[3], MD; Mohaned Shilaih[2], DPhil; Brigitte Leeners[1,4], MD

[1]Department of Reproductive Endocrinology, University Hospital Zurich, Zurich, Switzerland
[2]Ava AG, Zurich, Switzerland
[3]Department of Obstetrics, University Hospital Zurich, Zurich, Switzerland
[4]University of Zurich, Zurich, Switzerland

Corresponding Author:
Brigitte Leeners, MD
Department of Reproductive Endocrinology
University Hospital Zurich
Rämistrasse 100
Zurich, 8091
Switzerland
Phone: 41 44 225 50 01
Email: Brigitte.Leeners@usz.ch

Abstract

Background: As a daily point measurement, basal body temperature (BBT) might not be able to capture the temperature shift in the menstrual cycle because a single temperature measurement is present on the sliding scale of the circadian rhythm. Wrist skin temperature measured continuously during sleep has the potential to overcome this limitation.

Objective: This study compares the diagnostic accuracy of these two temperatures for detecting ovulation and to investigate the correlation and agreement between these two temperatures in describing thermal changes in menstrual cycles.

Methods: This prospective study included 193 cycles (170 ovulatory and 23 anovulatory) collected from 57 healthy women. Participants wore a wearable device (Ava Fertility Tracker bracelet 2.0) that continuously measured the wrist skin temperature during sleep. Daily BBT was measured orally and immediately upon waking up using a computerized fertility tracker with a digital thermometer (Lady-Comp). An at-home luteinizing hormone test was used as the reference standard for ovulation. The diagnostic accuracy of using at least one temperature shift detected by the two temperatures in detecting ovulation was evaluated. For ovulatory cycles, repeated measures correlation was used to examine the correlation between the two temperatures, and mixed effect models were used to determine the agreement between the two temperature curves at different menstrual phases.

Results: Wrist skin temperature was more sensitive than BBT (sensitivity 0.62 vs 0.23; *P*<.001) and had a higher true-positive rate (54.9% vs 20.2%) for detecting ovulation; however, it also had a higher false-positive rate (8.8% vs 3.6%), resulting in lower specificity (0.26 vs 0.70; *P*=.002). The probability that ovulation occurred when at least one temperature shift was detected was 86.2% for wrist skin temperature and 84.8% for BBT. Both temperatures had low negative predictive values (8.8% for wrist skin temperature and 10.9% for BBT). Significant positive correlation between the two temperatures was only found in the follicular phase (*rmcorr* correlation coefficient=0.294; *P*=.001). Both temperatures increased during the postovulatory phase with a greater increase in the wrist skin temperature (range of increase: 0.50 °C vs 0.20 °C). During the menstrual phase, the wrist skin temperature exhibited a greater and more rapid decrease (from 36.13 °C to 35.80 °C) than BBT (from 36.31 °C to 36.27 °C). During the preovulatory phase, there were minimal changes in both temperatures and small variations in the estimated daily difference between the two temperatures, indicating an agreement between the two curves.

Conclusions: For women interested in maximizing the chances of pregnancy, wrist skin temperature continuously measured during sleep is more sensitive than BBT for detecting ovulation. The difference in the diagnostic accuracy of these methods was likely attributed to the greater temperature increase in the postovulatory phase and greater temperature decrease during the menstrual phase for the wrist skin temperatures.

(J Med Internet Res 2021;23(6):e20710) doi:10.2196/20710

KEYWORDS

ovulation; basal body temperature; BBT; oral temperature; wrist skin temperature; diagnostic accuracy; thermometer; fertility; menstruation; wearable; sensor; mobile phone

Introduction

Background

Basal body temperature (BBT) is the lowest body temperature in the circadian rhythm. Monitoring BBT is one of the simplest and least invasive methods for determining the occurrence of ovulation and estimating its timing during the menstrual cycle [1]. In most women, BBT reaches its lowest point in each cycle just before ovulation and increases in the luteal phase because of the thermogenic properties of progesterone [2]. This physiological change is known as a temperature shift that presents as a biphasic pattern on the BBT curve [3].

Oral temperature taken immediately upon waking is widely used for measuring BBT by women who are interested in tracking their menstrual cycles or women who are trying to conceive because it is easy to use and noninvasive [4]. As a daily point measurement, BBT curves are sensitive to missing values and the time of measurement; a temperature shift may go undetected because a rise in body temperature may not have occurred yet at the time of the measurement [5]. In addition, lifestyle factors may strongly influence the reliability of this method [6]. A number of devices offering continuous temperature measurements at different body sites have been developed over the past decade [7-13]. In our previous studies, we have shown that continuously measured wrist skin temperature during sleep also presented a biphasic pattern in menstrual cycles, with 82% of the observed cycles having a sustained 3-day temperature shift [9,12].

Objectives

The primary objective of this study is to determine whether continuously measured wrist skin temperature during sleep was more accurate in detecting ovulation than BBT measured by oral temperature, using luteinizing hormone (LH) tests as the standard reference. The secondary objective is to investigate the correlation and agreement between these two temperatures in describing thermal changes in menstrual cycles.

Methods

Study Design and Participants

This prospective comparative diagnostic accuracy study was conducted from February to August 2019. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Cantonal Ethics Committee of Zurich, Switzerland (BASEC-Nr 2016-02241). All participants provided written informed consent before any study procedures were performed.

A convenience sample of participants was recruited through social media advertisements and networks from January to February 2019. At the time of enrollment, the research staff assessed the eligibility of potential participants using a screening questionnaire. This assessment was conducted at the Department of Reproductive Endocrinology, University Hospital of Zurich. Healthy women who met the following criteria were considered eligible: aged 18-45 years, not currently on hormonal therapy, willing to comply with the study protocol for up to six cycles, not planning pregnancy within the following 6 months, and currently living in Switzerland. Women were excluded if they had problems wearing the bracelet, had difficulty understanding the study procedures, had any health-related issues potentially affecting their menstrual cycles, were taking any medication or other substances that could affect the menstrual cycles or any physiological parameters being studied, were working night shifts or frequently traveling between different time zones, had a sleeping disorder or slept less than 4 hours per night, or were actively breastfeeding. Eligibility criteria had no restrictions on the regularity or length of menstrual cycle.

Information on age, body weight, body height, race, and time since stopping hormonal contraception was collected after receiving informed consent. BMI was calculated as kg/m^2. During the study period, participants measured continuous wrist skin temperature using the Ava Fertility Tracker, BBT using the Lady-Comp, and a home-based urine LH test using the ClearBlue Digital Ovulation Test (Swiss Precision Diagnostics GmbH). Participants received all the study materials, including detailed guidelines and study devices. The research staff provided participants with instructions on using the study devices and steps to be completed during the study. Contact details of the research staff and technical support staff for the Ava Fertility Tracker are provided. Participants were instructed to start all study measurements from the first day of enrollment, independent of their menstrual cycle day. This was done to enhance compliance and ensure that any issues were promptly resolved before the commencement of the next cycle.

Continuous Wrist Skin Temperature and BBT Measurement

Participants wore the Ava Fertility Tracker bracelet (version 2.0, Ava AG) on the dorsal side of their wrist (always of the same arm) each night while sleeping. The bracelet measures four physiological parameters simultaneously: wrist skin temperature, heart rate, heart rate variability, and breathing rate. The latter three parameters were not of interest to this study and thus were not included. At least 4 hours of relatively uninterrupted sleep each night is required for the physiological parameters to stabilize according to the manufacturer's packaging. The bracelet automatically saves physiological information every 10 seconds throughout the night. For this study, the first 90 minutes and the last 30 minutes of each night's data were excluded to avoid disturbances of the falling asleep and waking up phases. Temperature data were smoothed using locally weighted scatterplot smoothing. The 99th percentile (stable maxima) was chosen out of several percentiles (10th, 50th, and 90th percentiles) as the daily wrist skin temperature in the final analyses [12]. During the initial interview, participants were shown how to synchronize the device with the complementary Ava app on their smartphones and were instructed to synchronize their data each morning after waking up. During synchronization, the anonymized device data were transferred to the server. After completion of the study, the research staff retrieved the wrist skin temperature data obtained during the study period from the server for the final analysis.

BBT was measured by Lady-Comp (Valley Electronics AG), which is a computerized fertility tracking device with a digital thermometer. The device includes a display panel that provides immediate temperature readings to its user. The participants measured their oral temperature using the device each morning immediately after waking up, before getting out of bed, and before starting any activity such as drinking water. BBT data were retrieved by connecting the device to a computer. For each participant, a file containing the BBT data obtained during the study period was downloaded from the manufacturer's website. After completing the study, participants either retrieved the data and sent them to the research staff or they sent the device to the research staff, who retrieved the data. The device was returned to the participants after the data retrieval.

Participants recorded the first day of menstrual bleeding on both the Ava app and Lady-Comp. The first day of bleeding was defined as the first day of the cycle. In case of discrepancies, the date on the Ava app was used.

LH Test

Participants performed a home-based urine LH test using the ClearBlue Digital Ovulation Test [14] for each cycle according to the manufacturer's instructions. Home-based LH tests are widely used to detect ovulation and determine the fertile window [1,15]. For this study, the test was performed on prespecified days of the participant's menstrual cycle. The starting day was calculated by identifying a participant's average number of cycle days and then subtracting it by 17 days. After the initial test, the participant continued doing the LH test daily until a positive result, which was shown as *a stable smiling face* symbol on the device, or until the next menstruation. A positive result indicates an LH surge, which typically occurs 1 day before ovulation [13]. Participants reported the LH test results in the dedicated field of their Ava app. The day following the LH surge was defined as the day of ovulation. A cycle with a positive LH test was considered as an ovulatory cycle, and one with only negative LH tests was considered an anovulatory cycle. The LH test served as the reference standard for evaluating the diagnostic accuracy of the two temperatures.

Statistical Analyses

This study was a subanalysis of a prospective diagnostic accuracy study that compared the 2 fertility tracking devices with the LH test. The hypothesis of the main study was that the Ava Fertility Tracker bracelet was equivalent to the Lady-Comp and LH tests in determining the ovulation day. The final analyses in the main study were restricted to ovulatory cycles. Assuming a clinically meaningful margin of ±2 days, SD of 3 days, a mean difference of 0, an intraclass correlation coefficient of 0.147, and three cycles per woman, 39 cycles from 13 women were required with 90% power and a one-sided α of .025. Assuming a 20% loss to follow-up and up to 50% of the cycles were excluded because of being anovulatory or missing data, a total of 58 women were planned for this study. This subanalysis study only used the temperatures collected by the 2 devices as index tests. All ovulatory and anovulatory cycles, except those affected by missing data, were included in this subanalysis study.

A simplified diagram presented the numbers of screened, eligible, consented, compliant, withdrawing, and lost to follow-up participants. Cycles with missing LH test results or ≥30% missing temperature measurements of either device were excluded from the final analysis. The baseline characteristics of the participants and their cycles were summarized using descriptive statistics. Continuous parameters were summarized as mean (SD), and categorical parameters were summarized as frequency (%). A temperature shift occurred if three temperature measurements were 0.2 °C above the highest value of the previous six measurements or of the previous five out of six measurements in case of missing values [1]. On the basis of this definition, multiple temperature shifts could be detected within a cycle. To avoid including variations in temperatures because of reasons other than the menstrual rhythm, such as sickness, we analyzed only those temperature shifts occurring during the last 14 days of a cycle. The total number of temperature shifts per cycle, the percentage of cycles showing at least one temperature shift, and the first day of the temperature shift relative to the day of ovulation were recorded. The diagnostic accuracy of using at least one temperature shift on the two temperature curves for detecting ovulation was evaluated using the LH test as the reference standard. Diagnostic accuracy measures included sensitivity, specificity, and predictive values, considering the nested design of the study.

The analyses of correlation and agreement between the two temperatures were performed only for ovulatory cycles because the ovulation day was required to separate the menstrual phases. Correlations between mean wrist skin temperature and mean BBT at the follicular and luteal phases were examined using repeated measures correlation (R package *rmcorr*). The *rmcorr* correlation coefficient (r_{rm}) determines the common intraindividual association for paired measurements assessed on two or more occasions for multiple individuals [16]. Linear mixed effects models with random intercepts and random slopes were used to examine the agreement between the curves of the wrist skin temperatures and BBT. In these analyses, daily temperature measurements were nested within cycles, and the cycles were nested within the participants. Linear mixed effects models allow the modeling of repeated measurements, further accounting for correlated intraindividual and intracycle observations [12]. Daily temperature measurements were estimated from the models. The temperature curves at the cycle level were smoothed using locally weighted scatterplot smoothing. As curve patterns changed within a cycle, the agreement between curves was examined in three separate phases: (1) the menstrual phase referred to the period from day 1 to day 5 of a menstrual cycle; (2) the preovulatory phase referred to the period starting 10 days before ovulation and lasting through the day of ovulation; and (3) the postovulatory phase referred to the period starting 1 day after ovulation to 10 days after ovulation. A temperature curve was considered biphasic if at least one temperature shift was present. A curve was monophasic if a temperature shift was absent.

All statistical analyses were performed using the R software (version 3.6.0). All hypotheses were two-tailed. Statistical significance was set at *P*<.05.

Results

Participants and Cycles

In total, 266 cycles were collected from 63 women (Figure 1). The final analysis included 193 cycles of 57 women. A total of 88.1% (170/193) of cycles from 55 women were ovulatory (cycles with positive LH test), and 11.9% (23/193) from 18 women were anovulatory (cycles with only negative LH test). Table 1 presents the characteristics of the participants and their cycles. In total, 9% (2/23) anovulatory cycles were less than 24 days, and 13% (3/23) were more than 35 days. No ovulatory cycles were less than 24 days, but 4.1% (7/170) were more than 35 days. Most of the ovulatory cycles (146/170, 85.9%) had a luteal phase length between 11 days and 17 days and only 13.5% (23/170) were ≤10 days.

Figure 1. Study flowchart. BBT: basal body temperature; LH: luteinizing hormone.

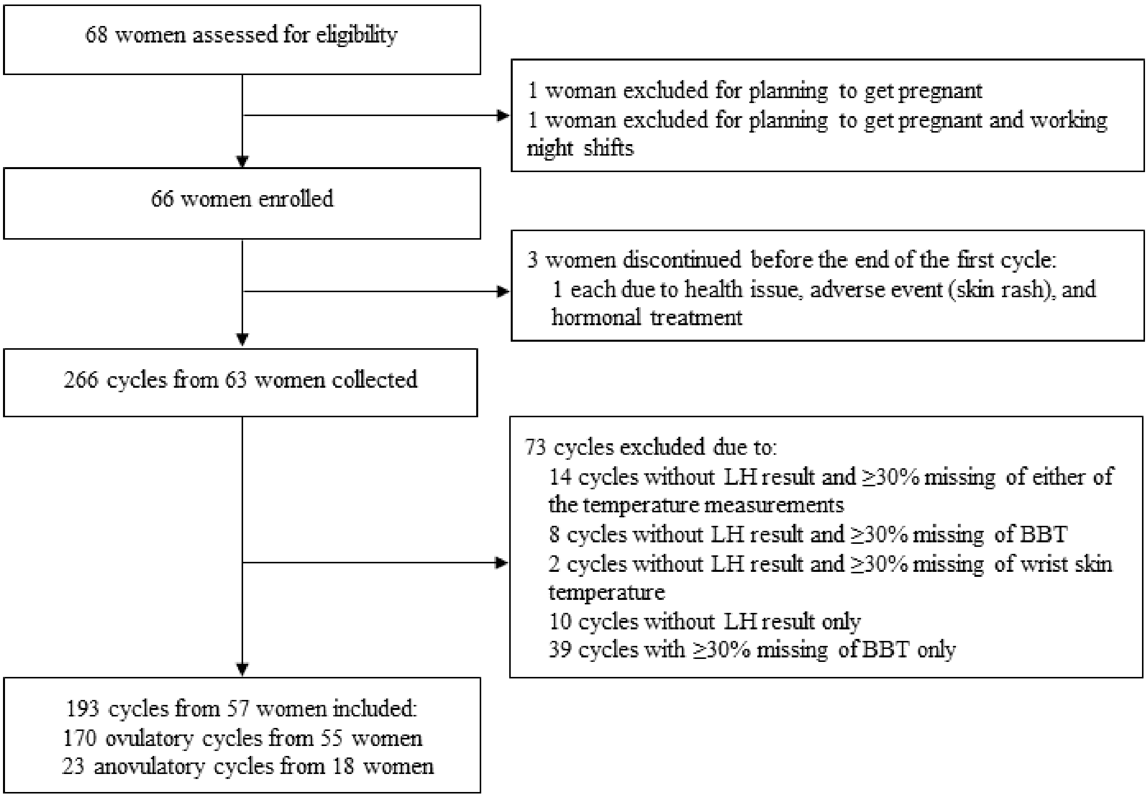


Table 1. Characteristics of the participants and their cycles.

Characteristics	Results
Participant level (n=57)	
Age (years), mean (SD)	26.7 (4.2)
Age (years), min-max	18-35
Age groups (years), n (%)	
18-20	5 (9)
21-25	19 (33)
26-30	23 (40)
31-35	10 (18)
Height (cm), mean (SD)	166.5 (6.0)
Weight (kg), mean (SD)	62.4 (9.9)
BMI (kg/m^2), mean (SD)	22.5 (3.6)
Race, n (%)	
White	43 (75)
Asian	3 (5)
Others	11 (19)
Time since stopping hormonal contraception (month), n (%)	
≥3	4 (7)
4-6	4 (7)
7-9	2 (4)
10-12	16 (28)
>12	17 (30)
No answer	14 (25)
Number of cycles per woman, mean (SD)	3.5 (1.0)
Women with 6 cycles, n (%)	1 (2)
Women with 5 cycles, n (%)	9 (16)
Women with 4 cycles, n (%)	17 (30)
Women with 3 cycles, n (%)	23 (40)
Women with 2 cycles, n (%)	5 (9)
Women with 1 cycle, n (%)	2 (4)
Cycle level (n=193)	
Cycle length (days), mean (SD)	29.5 (4.5)
Cycle length (days), min-max	21-60
Anovulatory cycles, n (%)	23 (11.9)
Cycle length (days), min-max	21-57
Ovulatory cycles, n (%)	170 (88.1)
Cycle length (days), min-max	24-60
Luteal length (days), mean (SD)	12.2 (1.9)
Luteal length, min-max	3-20

Temperature Shifts and Diagnostic Accuracy

For ovulatory cycles, the percentage of cycles with at least one temperature shift was significantly higher on the wrist skin temperature curves than that on the BBT curves (106/170, 62.4% vs 39/170, 22.9%; *P*<.001); however, the temperature shift occurred almost 2 days earlier on the BBT curves than on the wrist skin temperature curves (*P*<.001; Table 2). For anovulatory cycles, the percentage of cycles with at least one temperature shift was also significantly higher on the wrist skin temperature than on the BBT curves (17/23, 74% vs 7/23, 30%; *P*=.004).

Using the LH test as the reference standard for ovulation, the wrist skin temperature was more sensitive than the BBT (sensitivity 0.62 vs 0.23; *P*<.001) with a higher true-positive rate (106/193, 54.9% vs 39/193, 20.2%); however, it also had a higher false-positive rate (17/193, 8.8% vs 7/193, 3.6%), which resulted in a lower specificity (0.26 vs 0.70; *P*=.002; Table 2).

The positive predictive value was slightly higher for the wrist skin temperature. For a temperature shift detected on a wrist skin temperature curve, there was an 86.2% probability of ovulation. On the BBT curve, this probability was 84.8%. The negative predictive value was low for both temperatures (*P*=.39). If no temperature shift was shown on a wrist skin temperature curve, there was only an 8.6% probability that this had been an anovulatory cycle. This probability on a BBT curve was slightly higher (10.9%), but the difference was not statistically significant (*P*=.74).

Table 2. Temperature shifts and diagnostic accuracy for wrist skin temperature and basal body temperature.

Variables	Wrist skin temperature	Basal body temperature
Ovulatory cycles		
Total number of temperature shifts detected, n	240	47
Cycles with ≥1 temperature shift, n (%)	106 (62.4)[a]	39 (22.9)[a]
With 1 temperature shift	40 (37.7)[b]	31 (79)[c]
With 2 temperature shifts	31 (29.2)[b]	8 (21)[c]
With 3 temperature shifts	15 (14.2)[b]	0 (0)
With >3 temperature shifts	20 (18.9)[b]	0 (0)
The first day of temperature shift relative to ovulation day, mean (SD)	4.4 (2.75)	2.69 (1.89)
Anovulatory cycles		
Total number of temperature shifts detected, n	39	9
Cycles with ≥1 temperature shift, n (%)	17 (74)[d]	7 (30)[d]
With 1 temperature shift	5 (29)[e]	5 (71)[f]
With 2 temperature shifts	4 (24)[e]	2 (29)[f]
With 3 temperature shifts	6 (35)[e]	0 (0)
With 4 temperature shifts	2 (12)[e]	0 (0)
Diagnostic accuracy (urine luteinizing hormone tests as standard reference; N=193)		
True positives, n (%)	106 (54.9)	39 (20.2)
True negatives, n (%)	6 (3.1)	16 (8.3)
False positives, n (%)	17 (8.8)	7 (3.6)
False negatives, n (%)	64 (33.2)	131 (67.9)
Sensitivity (95% CI)	0.62 (0.55-0.70)	0.23 (0.17-0.30)[g]
Specificity (95% CI)	0.26 (0.10-0.48)	0.70 (0.47-0.87)[g]
Negative predictive value (95% CI)	0.09 (0.03-0.18)	0.11 (0.06-0.17)[g]
Positive predictive value (95% CI)	0.86 (0.79-0.92)	0.85 (0.71-0.94)[g]

[a] N=170.

[b] N=106.

[c] N=39.

[d] N=23.

[e] N=17.

[f] N=7.

[g] *P* values comparing wrist skin temperature and basal body temperature: *P*<.001 for sensitivity; *P*=.002 for specificity; *P*=.39 for negative predictive value; and *P*=.74 for positive predictive value.

Correlation of the Two Temperatures

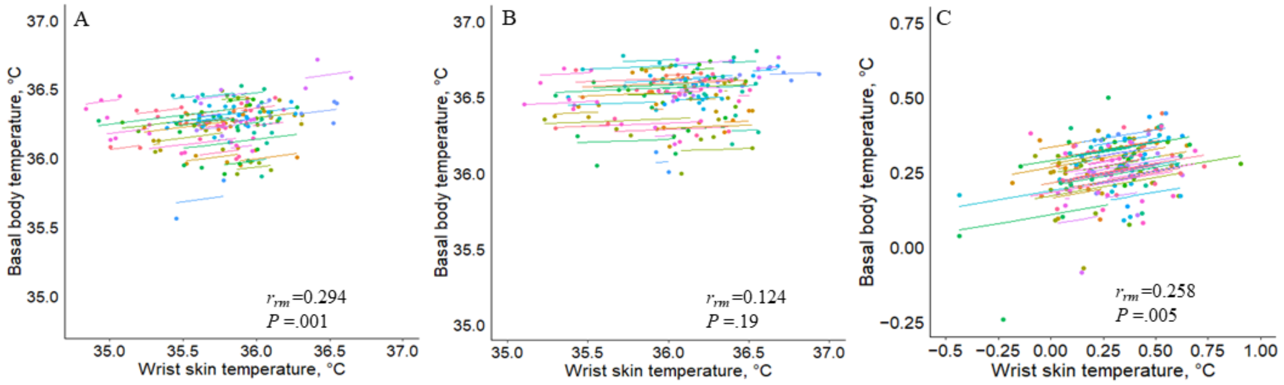
Both temperatures differed between the follicular and luteal phases, with the latter having higher temperatures (Table 3). Throughout the menstrual cycle, the wrist skin temperature was generally lower than BBT. The mean between-phase temperature change was 11% higher for the wrist skin temperature than for the BBT.

Table 3. Mean follicular and luteal phase temperatures for wrist skin temperature and basal body temperature.

Phases	Wrist skin temperature (°C), mean (SD)	Basal body temperature (°C), mean (SD)
Follicular phase	35.78 (0.34)	36.25 (0.16)
Luteal phase	36.07 (0.35)	36.51 (0.16)
Between-phase temperature change	0.29 (0.21)	0.26 (0.1)

Figure 2 shows the repeated measures correlation plot for the means of the two temperatures. Each dot represents the mean of the two temperatures in one menstrual cycle of a woman. Observations from the same woman are given the same color, with corresponding lines showing the repeated measures correlation fit for each woman. Positive correlations were found only in the follicular phase (r_{rm}=0.294; 95% CI 0.117-0.454; *P*=.001). This correlation showed a minimal interindividual variation, which was reflected by the mostly parallel lines. In the luteal phase, no correlation was found between the two temperatures (r_{rm}=0.124; 95% CI −0.061 to 0.301; *P*=.19). We found positive correlations for between-phase temperature changes measured at the two temperatures (r_{rm}=0.258; 95% CI 0.078-0.422; *P*=.005).

Figure 2. Repeated measures correlation (rmcorr) plots of mean wrist skin temperature and basal body temperature. A: correlation in the follicular phase; B: correlation in the luteal phase; C: correlation of between-phase temperature changes between the two temperatures.



Agreement of Temperature Curves

The agreement of temperature curves was analyzed for ovulatory cycles. Figure 3 shows the smoothed curves of wrist skin temperatures and BBT during the postovulatory, menstrual, and preovulatory phases. There was no overlap between the two curves, and the agreement differed by phase. Changes in the two temperatures were not observed during the postovulatory and menstrual phases. Both temperature values increased during the postovulatory phase with a greater and more continuous increase in the wrist skin temperature (range of increase 0.50 °C vs 0.20 °C). The estimated daily difference between the two temperatures was the greatest on day 2 (0.64 °C) and the smallest on day 10 after ovulation (0.32 °C), with a mean of 0.49 °C (*P*<.001). During the menstrual phase, the wrist skin temperature exhibited a more substantial decrease (from 36.13 °C to 35.80 °C, range of decrease: 0.33 °C) than BBT (from 36.31 °C to 36.27 °C; range of decrease: 0.04 °C). The estimated daily difference between the two temperatures ranged from 0.18 °C on day 1 to 0.46 °C on day 5, with a mean difference of 0.32 °C (*P*<.001). During the preovulatory phase, there were minimal changes in both the wrist skin temperature (range: 0.09 °C) and BBT (range: 0.07 °C) and a small variation in the estimated daily difference between the two temperature values, indicating an agreement between the two curves. The mean daily difference between the two temperatures was 0.53 °C (*P*<.001).

Figure 3. Smoothed temperature curves according to phases. A: postovulatory phase; B: menstrual phase; C: preovulatory phase. Solid and dotted lines represent wrist skin temperature and 95% CIs, respectively; dashed and dotted lines represent basal body temperature and 95% CIs, respectively.

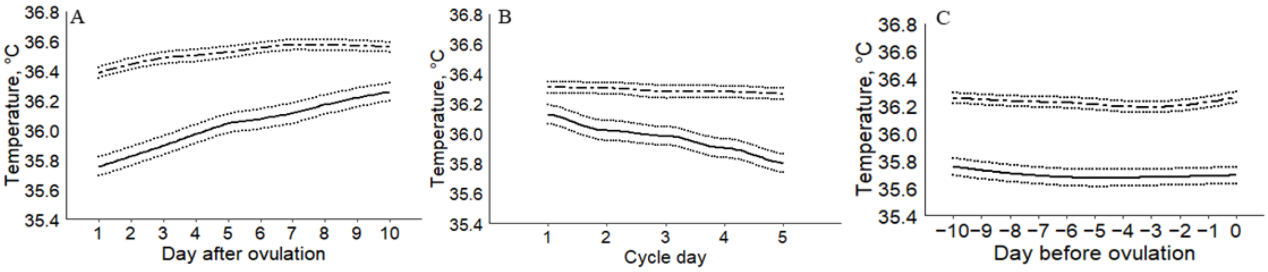
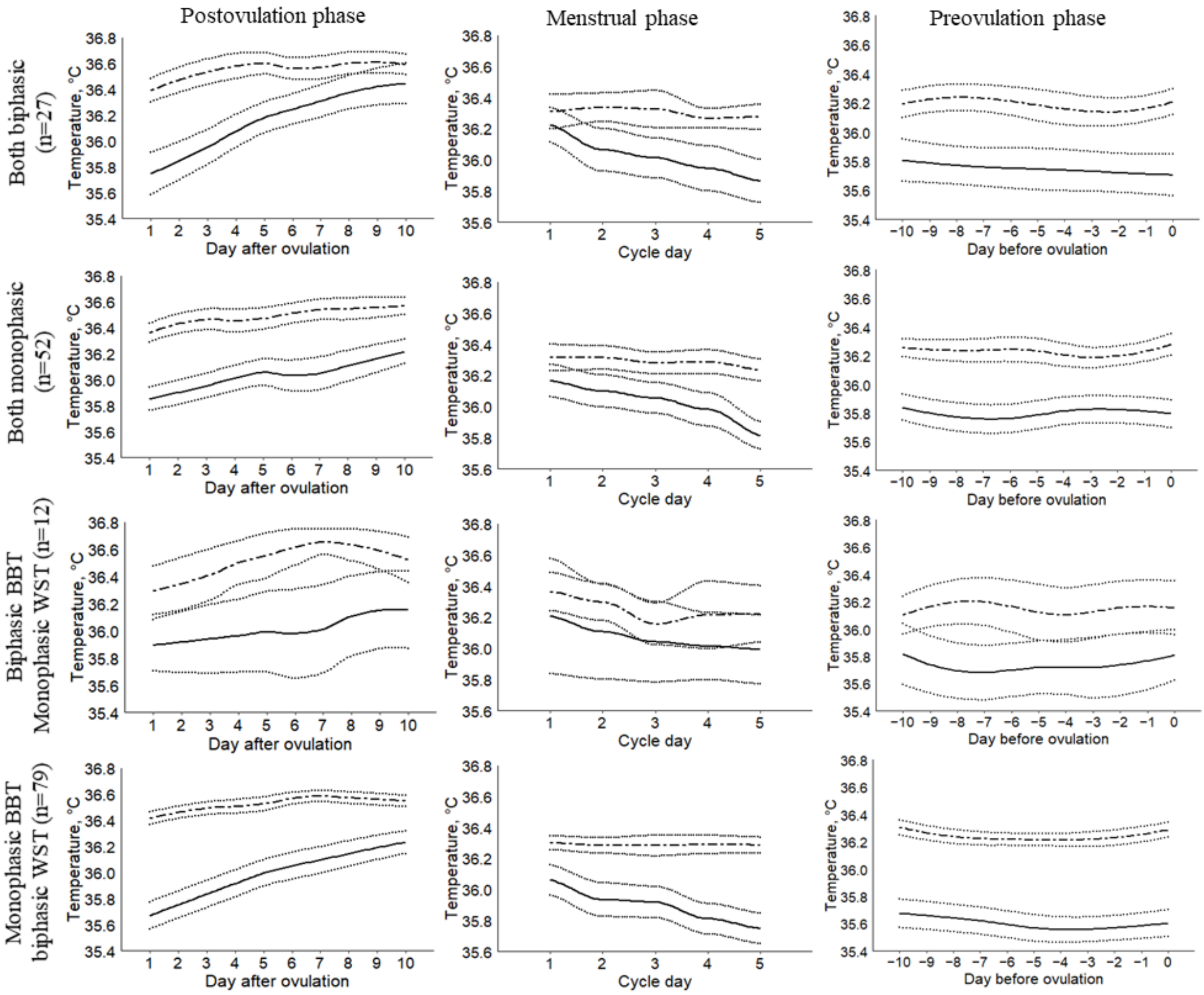


Figure 4 shows the agreement based on the curve patterns. A significant overlap of the two curves was found when the BBT curve was biphasic and the wrist skin temperature curve was monophasic. However, the wide CIs could be the result of the small number of cycles in this category (n=12). The agreement between the two curves, particularly during the postovulatory phase, was highest when both curves were monophasic (n=52), with estimated daily differences between the two temperatures ranging from 0.34 °C to 0.57 °C. The most prominent disagreement during the postovulatory phase was observed when the wrist skin temperature curve was biphasic and the BBT curve monophasic (n=79), with daily differences between the two temperatures ranging from 0.32 °C to 0.74 °C. Even when both curves were biphasic (n=27), the disagreement could still be observed during the early postovulatory phase, with the largest difference between the two temperatures on day 2 after ovulation (0.67 °C). The two curves overlapped during the late postovulatory phase.

Figure 4. Smoothed temperature curves according to phases and patterns. Solid and dotted lines represent wrist skin temperature and 95% CIs, respectively; dashed and dotted lines represent basal body temperature and 95% CIs, respectively. BBT: basal body temperature; WST: wrist skin temperature.



Discussion

Principal Findings

Continuously measured wrist skin temperature had a higher sensitivity and lower specificity for detecting ovulation than BBT measured orally. The two temperatures were significantly correlated in the follicular phase but not in the luteal phase, indicating that changes in the 2 temperatures were not coupled in the luteal phase. The wrist skin temperature curve showed a greater increase during the postovulatory phase and a greater decrease in the menstrual phase than the BBT curve. The disagreement between the temperature curves was most prominent when the BBT curve was monophasic and the wrist skin temperature curve was biphasic. Our results suggest that the continuously measured wrist skin temperature is more sensitive than BBT to detect ovulation and determine the fertile window.

This study is the first to compare the wrist skin temperature and BBT in detecting ovulation. As skin tissues are not close to major blood vessels but exposed to the environment and heat transfer from the core to surface tissue is not instantaneous, the skin temperature is generally lower than BBT but not by a fixed amount [17]. Previous studies have explored how continuously measured temperatures at various body sites change across different menstrual phases. The study by Maijala et al [10] showed that both finger skin temperature measured nocturnally and oral temperature measured upon waking up differed between the luteal and follicular phases. The average between-phase temperature change was 0.07°C higher for the finger skin temperature, and this difference was statistically significant [10]. The intestinal core temperature showed changes in the circadian rhythm over the menstrual cycle, with a higher daily mean temperature during the luteal phase than during the follicular phase, and the daily minimum temperature was at its lowest value before the LH peak [14]. Regidor et al [11] reported an accuracy of 88.8% in predicting a window of 3 days before ovulation, the day of ovulation, and 3 days after ovulation with a device that continuously measured the vaginal temperature throughout the menstrual cycle. Details about the device or the methods used to evaluate the temperature curves were not provided in their paper. Findings from our previous research confirmed that wrist skin temperature could detect menstrual phase–related temperature changes, and these changes were robust to common confounding factors known to affect BBT such as sexual activity, food intake, and alcohol consumption [9,12].

One particularly encouraging finding of this study is the higher sensitivity of the continuous wrist skin temperature than oral BBT, which indicates that continuous rather than point measurement may be more useful for pregnancy planning. Moreover, the nightly data of wrist skin temperature were handled in a more sophisticated mathematical way, allowing it to be more sensitive in detecting temperature shift in a menstrual cycle. Its lower specificity may reflect a similar trade-off between the sensitivity and specificity of any diagnostic test. One compelling argument is that specificity is more relevant for avoiding an unplanned pregnancy, as the test should have both a high true negative rate and a high negative predictive value. Although we can conclude that wrist skin temperature is more sensitive than BBT for maximizing the chances of pregnancy, neither temperature should be used as a standalone method to avoid an unplanned pregnancy given their low negative predictive values.

The difference in diagnostic accuracy between the two temperatures can be explained by the different thermal changes revealed by the two temperature curves. In particular, compared with BBT, the wrist skin temperature exhibited a greater magnitude of increase during the postovulatory phase. This allows the wrist skin temperature to be more sensitive in detecting the temperature shift. In addition, the wrist skin temperature increased in a steeper and more continuous manner. This explains why it detected more temperature shifts than BBT. Few studies have compared the two temperature curves [4,13]. Krauchi et al [4] reported a similar pattern of temperature changes in a menstrual cycle between BBT and skin temperature. In contrast, Wark et al [13] reported poor agreement between BBT and the mean upper arm skin temperature measured at 10-, 30-, 60-, 90-, and 120-minute intervals before waking.

The different modes of measurement and the different circadian rhythms of wrist skin and oral temperature, which are more pronounced during sleep [18-20], might explain the disagreement between the two temperatures. The oral temperature decreases continuously during sleep, with the lowest value occurring at approximately 5 AM, and then rises sharply after waking up [21]. Point measurements such as BBT are susceptible to variations in waking times and compliance because a single measurement is located on the sliding scale of the circadian rhythm. During the preovulatory period, the amplitude of the circadian rhythm reaches the highest value [14], which might further limit the ability of BBT to identify a temperature shift. In contrast, the circadian rhythm of wrist skin temperature features a sharp increase before lights off, a plateau at a higher temperature during sleep, and then a sharp drop immediately after rising [18,20]. Sleep propensity is accompanied by an increased skin blood flow and less cold-induced vasoconstriction, particularly in the distal skin areas that are most strongly involved in the regulation of heat loss because of their richness in arteriovenous anastomoses, thus increasing the skin temperature [22,23]. Once awake, cold-induced vasoconstriction is restored, thereby decreasing the skin temperature. In this study, participants wore a wearable device that continuously measured the wrist skin temperature during sleep. As the first 90 and the last 30 minutes of recorded data were excluded, the nocturnal steady state when the temperature was maintained at a high level was captured. As a result, the effect of the circadian rhythm was, to a certain extent, removed and the temperature changes reflected mostly the menstrual rhythm. Consequently, this measurement is more sensitive for detecting temperature shifts during menstrual cycles.

Limitations

Our study has several limitations. First, factors that could potentially influence temperature were not evaluated. These factors include sexual activity, exercise, food intake, alcohol consumption, sleep duration and quality, and wake-up time [6]. An examination of the potential influence of these factors on the curve patterns and their agreement would have been particularly interesting. Second, because of the relatively small number of participants, we were unable to conduct subgroup analyses on the influence of BMI or cycle length on the difference in diagnostic accuracy between the two temperatures. Third, 27.4% (73/266) of the collected cycles were excluded from the final analysis because of missing measurements. The percentage of cycles with ≥30% missing wrist skin temperature was comparable with that reported in a previous study [12]. It remains unclear whether compliance would improve under real-world conditions where users are actively track their menstrual cycles. Furthermore, the participants in our study were recruited by nonrandom sampling and consisted of mostly young White women. Novel digital technology might be particularly appealing to these participants. Whether our results are generalizable to other races and real-world conditions requires further study.

Conclusions

For women interested in maximizing the chances of pregnancy, the wrist skin temperature continuously measured during sleep is more sensitive than BBT to detect ovulation. The difference in the diagnostic accuracy of these two methods was likely attributed to the greater temperature increase in the postovulatory phase and a greater decrease during the menstrual phase for the wrist skin temperature. However, when used as a standalone method, neither of the temperatures could reliably avoid unplanned pregnancy, given the low negative predictive values. Our results underpin the importance of validation studies, especially against a standard reference test, while developing wearable devices that measure physiological parameters for women or clinicians to track menstrual cycles.

Acknowledgments

The authors gratefully acknowledge Innosuisse (17324.1 PFLS-LS, 28715.1 IP-LS, 40347.1 IP-LS) for funding the development of the wearable device. The authors thank all participating women for supporting our study.

Authors' Contributions

MR, GH, MS, and BL designed the study. GH and AH acquired the data. MR conducted the data analyses. TYZ and MR interpreted the data. TYZ drafted the manuscript. The draft was critically revised by TYZ, MR, JW, MK, NK, MS, and BL. All authors approved the final version of the manuscript for publication.

Conflicts of Interest

MR is an employee of Ava AG (Zurich, Switzerland). MS is a former employee of Ava AG. BL is a member of the advisory board of Ava AG.

References

1. Su H, Yi Y, Wei T, Chang T, Cheng C. Detection of ovulation, a review of currently available methods. Bioeng Transl Med 2017 May 16;2(3):238-246 [FREE Full text] [doi: 10.1002/btm2.10058] [Medline: 29313033]
2. Mesen TB, Young SL. Progesterone and the luteal phase: a requisite to reproduction. Obstet Gynecol Clin North Am 2015 Mar;42(1):135-151 [FREE Full text] [doi: 10.1016/j.ogc.2014.10.003] [Medline: 25681845]
3. Marshall J. A field trial of the basal-body-temperature method of regulating births. Lancet 1968 Jul 06;2(7558):8-10. [doi: 10.1016/s0140-6736(68)92886-9] [Medline: 4172715]
4. Kräuchi K, Konieczka K, Roescheisen-Weich C, Gompper B, Hauenstein D, Schoetzau A, et al. Diurnal and menstrual cycles in body temperature are regulated differently: a 28-day ambulatory study in healthy women with thermal discomfort of cold extremities and controls. Chronobiol Int 2014 Feb;31(1):102-113. [doi: 10.3109/07420528.2013.829482] [Medline: 24131147]
5. Bauman JE. Basal body temperature: unreliable method of ovulation detection. Fertil Steril 1981 Dec;36(6):729-733. [doi: 10.1016/s0015-0282(16)45916-9] [Medline: 7308516]
6. Barron ML, Fehring RJ. Basal body temperature assessment: is it useful to couples seeking pregnancy? MCN Am J Matern Child Nurs 2005;30(5):290-297. [doi: 10.1097/00005721-200509000-00004] [Medline: 16132004]
7. Berglund SE, Lindén HA, Scherwitzl R. Identification and prediction of the fertile window using NaturalCycles. Eur J Contracept Reprod Health Care 2015;20(5):403-408. [doi: 10.3109/13625187.2014.988210] [Medline: 25592280]
8. Demiańczyk A, Michaluk K. Evaluation of the effectiveness of selected natural fertility symptoms used for contraception: estimation of the Pearl index of Lady-Comp, Pearly and Daysy cycle computers based on 10 years of observation in the Polish market. Ginekol Pol 2016;87(12):793-797 [FREE Full text] [doi: 10.5603/GP.2016.0090] [Medline: 28098936]
9. Goodale BM, Shilaih M, Falco L, Dammeier F, Hamvas G, Leeners B. Wearable sensors reveal menses-driven changes in physiology and enable prediction of the fertile window: observational study. J Med Internet Res 2019 Apr 18;21(4):e13404 [FREE Full text] [doi: 10.2196/13404] [Medline: 30998226]

10. Maijala A, Kinnunen H, Koskimäki H, Jämsä T, Kangas M. Nocturnal finger skin temperature in menstrual cycle tracking: ambulatory pilot study using a wearable Oura ring. BMC Womens Health 2019 Nov 29;19(1):150 [FREE Full text] [doi: 10.1186/s12905-019-0844-9] [Medline: 31783840]
11. Regidor P, Kaczmarczyk M, Schiweck E, Goeckenjan-Festag M, Alexander H. Identification and prediction of the fertile window with a new web-based medical device using a vaginal biosensor for measuring the circadian and circamensual core body temperature. Gynecol Endocrinol 2018 Mar;34(3):256-260. [doi: 10.1080/09513590.2017.1390737] [Medline: 29082805]
12. Shilaih M, Goodale BM, Falco L, Kübler F, De Clerck V, Leeners B. Modern fertility awareness methods: wrist wearables capture the changes in temperature associated with the menstrual cycle. Biosci Rep 2018 Dec 21;38(6):A [FREE Full text] [doi: 10.1042/BSR20171279] [Medline: 29175999]
13. Wark JD, Henningham L, Gorelik A, Jayasinghe Y, Hartley S, Garland SM. Basal temperature measurement using a multi-sensor armband in australian young women: a comparative observational study. JMIR Mhealth Uhealth 2015 Oct 05;3(4):e94 [FREE Full text] [doi: 10.2196/mhealth.4263] [Medline: 26441468]
14. Coyne MD, Kesick CM, Doherty TJ, Kolka MA, Stephenson LA. Circadian rhythm changes in core temperature over the menstrual cycle: method for noninvasive monitoring. Am J Physiol Regul Integr Comp Physiol 2000 Oct;279(4):1316-1320 [FREE Full text] [doi: 10.1152/ajpregu.2000.279.4.R1316] [Medline: 11003999]
15. Behre HM, Kuhlage J, Gassner C, Sonntag B, Schem C, Schneider HP, et al. Prediction of ovulation by urinary hormone measurements with the home use ClearPlan Fertility Monitor: comparison with transvaginal ultrasound scans and serum hormone measurements. Hum Reprod 2000 Dec;15(12):2478-2482. [doi: 10.1093/humrep/15.12.2478] [Medline: 11098014]
16. Bakdash JZ, Marusich LR. Repeated measures correlation. Front Psychol 2017;8:456 [FREE Full text] [doi: 10.3389/fpsyg.2017.00456] [Medline: 28439244]
17. Baker FC, Siboza F, Fuller A. Temperature regulation in women: effects of the menstrual cycle. Temperature (Austin) 2020;7(3):226-262 [FREE Full text] [doi: 10.1080/23328940.2020.1735927] [Medline: 33123618]
18. Areas R, Duarte L, Menna-Barreto L. Comparative analysis of rhythmic parameters of the body temperature in humans measured with thermistors and digital thermometers. Biol Rhythm Res 2006 Oct;37(5):419-424. [doi: 10.1080/09291010600869752]
19. Cuesta M, Boudreau P, Cermakian N, Boivin DB. Skin temperature rhythms in humans respond to changes in the timing of sleep and light. J Biol Rhythms 2017 Jun;32(3):257-273. [doi: 10.1177/0748730417702974] [Medline: 28569119]
20. Sarabia JA, Rol MA, Mendiola P, Madrid JA. Circadian rhythm of wrist temperature in normal-living subjects A candidate of new index of the circadian system. Physiol Behav 2008 Nov 28;95(4):570-580. [doi: 10.1016/j.physbeh.2008.08.005] [Medline: 18761026]
21. Waterhouse J, Drust B, Weinert D, Edwards B, Gregson W, Atkinson G, et al. The circadian rhythm of core temperature: origin and some implications for exercise performance. Chronobiol Int 2005;22(2):207-225. [doi: 10.1081/cbi-200053477] [Medline: 16021839]
22. Kräuchi K, Wirz-Justice A. Circadian rhythm of heat production, heart rate, and skin and core temperature under unmasking conditions in men. Am J Physiol 1994 Sep;267(3 Pt 2):819-829. [doi: 10.1152/ajpregu.1994.267.3.R819] [Medline: 8092328]
23. Lindert BH, Van Someren EJ. Skin temperature, sleep, and vigilance. Handb Clin Neurol 2018;156:353-365. [doi: 10.1016/B978-0-444-63912-7.00021-7] [Medline: 30454600]

Abbreviations

BBT: basal body temperature
LH: luteinizing hormone

Edited by R Kukafka; submitted 06.10.20; peer-reviewed by D Bychkov, Y Motoki; comments to author 02.01.21; revised version received 15.01.21; accepted 19.04.21; published 08.06.21.

Please cite as:
Zhu TY, Rothenbühler M, Hamvas G, Hofmann A, Welter J, Kahr M, Kimmich N, Shilaih M, Leeners B
The Accuracy of Wrist Skin Temperature in Detecting Ovulation Compared to Basal Body Temperature: Prospective Comparative Diagnostic Accuracy Study
J Med Internet Res 2021;23(6):e20710
URL: https://www.jmir.org/2021/6/e20710
doi:10.2196/20710
PMID:34100763

Original Paper

Association Between the Digital Clock Drawing Test and Neuropsychological Test Performance: Large Community-Based Prospective Cohort (Framingham Heart Study)

Jing Yuan[1,2], MD; David J Libon[3], PhD; Cody Karjadi[4], MSc; Alvin F A Ang[2,4,5], MD, MPH; Sherral Devine[2,4], PhD; Sanford H Auerbach[4,6], MD; Rhoda Au[2,4,5,6,7*], MBA, PhD; Honghuang Lin[4,8*], PhD

[1]Department of Neurology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing, China

[2]Department of Anatomy and Neurobiology, School of Medicine, Boston University, Boston, MA, United States

[3]Department of Geriatrics and Gerontology and the Department of Psychology, New Jersey Institute for Successful Aging, School of Osteopathic Medicine, Rowan University, Stratford, NJ, United States

[4]Framingham Heart Study, School of Medicine, Boston University, Boston, MA, United States

[5]Slone Epidemiology Center, School of Medicine, Boston University, Boston, MA, United States

[6]Department of Neurology, School of Medicine, Boston University, Boston, MA, United States

[7]Department of Epidemiology, School of Public Health, Boston University, Boston, MA, United States

[8]Computational Biomedicine, School of Medicine, Boston University, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Honghuang Lin, PhD
Computational Biomedicine
School of Medicine
Boston University
72 E. Concord Street, E-632
Boston, MA, 02118
United States
Phone: 1 617 358 0091
Fax: 1 617 358 5677
Email: hhlin@bu.edu

Abstract

Background: The Clock Drawing Test (CDT) has been widely used in clinic for cognitive assessment. Recently, a digital Clock Drawing Text (dCDT) that is able to capture the entire sequence of clock drawing behaviors was introduced. While a variety of domain-specific features can be derived from the dCDT, it has not yet been evaluated in a large community-based population whether the features derived from the dCDT correlate with cognitive function.

Objective: We aimed to investigate the association between dCDT features and cognitive performance across multiple domains.

Methods: Participants from the Framingham Heart Study, a large community-based cohort with longitudinal cognitive surveillance, who did not have dementia were included. Participants were administered both the dCDT and a standard protocol of neuropsychological tests that measured a wide range of cognitive functions. A total of 105 features were derived from the dCDT, and their associations with 18 neuropsychological tests were assessed with linear regression models adjusted for age and sex. Associations between a composite score from dCDT features were also assessed for associations with each neuropsychological test and cognitive status (clinically diagnosed mild cognitive impairment compared to normal cognition).

Results: The study included 2062 participants (age: mean 62, SD 13 years, 51.6% women), among whom 36 were diagnosed with mild cognitive impairment. Each neuropsychological test was associated with an average of 50 dCDT features. The composite scores derived from dCDT features were significantly associated with both neuropsychological tests and mild cognitive impairment.

Conclusions: The dCDT can potentially be used as a tool for cognitive assessment in large community-based populations.

(J Med Internet Res 2021;23(6):e27407) doi:10.2196/27407

KEYWORDS

clock drawing test; neuropsychological test; cognition; technology; digital assessment; mild cognitive impairment; association; neurology; Framingham Heart Study

Introduction

The Clock Drawing Test (CDT) is a widely used neuropsychological test to screen cognitive impairment and dementia because of its ease of administration and clinical assessment capability [1,2]. The test is typically administered by specifying a time, for example, ten past eleven, and asking patients or participants to draw a clock showing that time (the command condition), followed by asking patients or participants to copy a predrawn clock image (the copy condition). Both test conditions require multiple cognitive domains. The command test condition requires intact attention, auditory comprehension, semantic memory, executive function, and visuoconstructional abilities, whereas the copy test condition relies primarily upon visuospatial, attention, and executive function skills [3-6]. Keen observation of the process by which drawings are produced is key to the evaluation of the type and severity of cognitive impairment [7,8]. Multiple manual scoring systems have been created to objectively quantify test performance. However, none of these scoring systems can capture the full breadth of cognitive skills used in completing the test [3].

Recently, a digital version of the CDT (dCDT) that uses a digital ballpoint pen and smart paper was developed as an alternative to the standard clock drawing scoring systems [9,10]. The digital pen can record its position with a timestamp and has excellent precision in capturing all graphomotor, spatial, and temporal information [11-13]; however, the characterization of these features and their correlations with standard neuropsychological tests has yet to be examined in a large community-based setting.

The objective of this investigation was to examine the association between dCDT features and cognitive functions in the Framingham Heart Study (FHS) cohort. We also investigated the association between dCDT features and cognitive status (clinically diagnosed mild cognitive impairment compared to those with normal cognition).

Methods

Study Sample

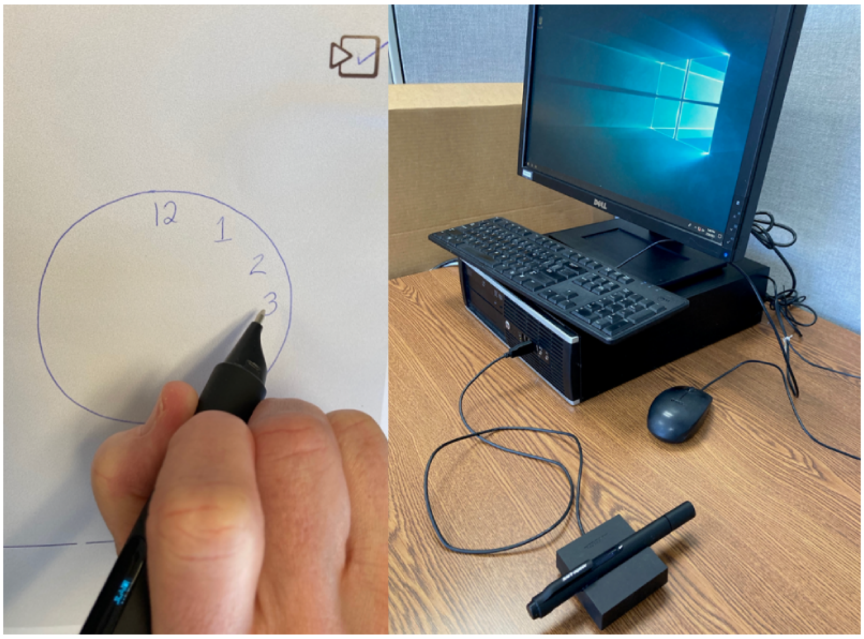
The FHS is a community-based prospective cohort study that was established in 1948. Three generations of participants have been enrolled. Details about the FHS cohort have been previously published [14-16]. This study included participants who completed at least one dCDT and neuropsychological assessment. Participants with prevalent dementia (n=23) or who had not been reviewed by the expert panel (n=138) were excluded. The Boston University Medical Campus Institutional Review Board approved the study procedures and protocols. Written informed consent was obtained from all participants.

The dCDT

Since October 2011, FHS participants who have come for their regular neuropsychological test visit were simultaneously administered the dCDT using a digital pen. The test was jointly developed by the Massachusetts Institute of Technology and the Lahey Hospital and Medical Center with the collaboration of the Clock Sketch Consortium [9-11,17]. Participants used the digital pen (Anoto Inc), to draw a clock on smart paper with a faint dot pattern (Figure 1). The digital pen functions as a regular ballpoint pen does but also measures the pen's position 80 times per second at a spatial resolution of 0.002 inches with a built-in camera [10,11]. Drawings are automatically classified into different categories, such as numbers, hands, and lines. For quality control, an external rater can replay and deconstruct each drawing to ensure appropriate classification. It typically takes 1 to 2 minutes to classify clocks drawn by healthy or mildly impaired individuals. Any classification errors can be corrected by the rater using a user-friendly drag-and-drop interface. Additional time may be required for classification of more complex clock images, however, most tasks are completed within 5 minutes.

More than 100 dCDT features have been derived to measure the entire drawing process, including capturing all strokes and their corresponding latencies. These features reflect a range of cognitive functions related to drawing efficiency, simple motor operations, information processing speed, and spatial reasoning (Multimedia Appendix 1, Table S1). The rank-based inverse normal transformation was later applied to all dCDT features to reduce the distribution skewness.

Figure 1. Digital Clock Drawing Test digital pen, smart paper, and docking setup.



Neuropsychological Assessment

A neuropsychological test protocol, measuring multiple cognitive domains of verbal memory, visual memory, attention and concentration, executive function, abstract reasoning, language, and visuoperceptual organization, was administered to all FHS participants [18]: Wechsler Memory Scale [19] Logical Memory—Immediate Recall, Delayed Recall, and Recognition; Visual Reproduction—Immediate Recall, Delayed Recall, and Recognition; Paired Associate Learning—Immediate Recall, Delayed Recall, and Recognition; the Wechsler Adult Intelligence Scale [20] Digit Span—Forward, Backward, and Similarities; Boston Naming Test—30-item version [21]; Trail Making Test A and B [22]; Hooper Visual Organization Test [23]; Verbal Fluency and Verbal Fluency—Animal [24,25]. All tests were administered by trained raters.

Ascertainment of Mild Cognitive Impairment

In addition to regular research center visits, FHS participants underwent neuropsychological assessments every 4 to 5 years [26,27]. For participants with possible cognitive impairment, regular neuropsychological tests were conducted every 1 to 2 years and neurological exams were performed on a subset of participants. When potential cognitive impairment decline was present, a clinical review was conducted by a panel with at least one neurologist and one neuropsychologist. Mild cognitive impairment diagnosis was determined by the review panel, which required that the patient exhibit evidence of a decline in cognitive performance in one or more cognitive domains, have no records indicating functional decline, and did not meet criteria for dementia. Although the Clinical Dementia Rating scale [28] was not formally applied, the panel used the Clinical Dementia Rating scoring scale (0-3) to quantify the severity of impairment; for mild cognitive impairment, a rating of 0.5 was given.

Statistical Analyses

Linear regression models were used to assess the association between each dCDT feature and neuropsychological tests. The models were adjusted for age and sex. Bonferroni correction was used to adjust for multiple testing, and significant associations were claimed if $P<.05/n_t$, where n_t was the number of tests performed; therefore, tests were significant if $P<4.8\text{x}10^{-4}$, with $\alpha=.05$ and 105 tests. A composite score was also created for each neuropsychological test based on dCDT features that were significantly associated with the test. The score for sample i is defined as

where m is the number of dCDT features significantly associated with the neuropsychological test, β_j is the estimate of effect size for feature j, and V_{ij} is the normalized dCDT feature j for sample i. The score represented a weighted combination of all dCDT features for the neuropsychological test. The associations between the composite score and each neuropsychological test were also tested with linear regression models adjusted for age and sex.

The association between neuropsychological tests and dCDT composite scores with mild cognitive impairment was assessed by logistic regression models adjusted for age, sex, and education. Age was treated as a continuous variable, whereas sex was treated as a dichotomous variable. Education was treated as a categorical variable (eg, no high school degree, high school degree, some college, and college graduate). The difference between groups was assessed with the Wilcoxon rank-sum test for continuous variables or the chi-square test for dichotomous and categorical variables. Bonferroni correction was used to adjust for multiple tests, and associations were significant if $P<2.8\times10^{-3}$, given that 18 neuropsychological tests were used. All statistical analyses were performed using R software (version 4.0.3, The R Project).

Results

As shown in Table 1, our study sample included 2,062 participants (age: mean 62, SD 13 years; 51.6% women; and 43.4% received college-level education or higher). Among them, 36 participants had been diagnosed with mild cognitive impairment. As expected, participants with mild cognitive impairment were generally older and had worse cognitive performance than those in the normal cognition group.

Table 1. Clinical characterization.

Variable	All (N=2062)	Mild cognitive impairment (n=36)	Normal cognition (n=2026)	*P* value
Age (years), mean (SD)	62 (13)	79 (7)	62 (13)	<.001
Gender, n (%)				.50
Women	1065 (51.6)	21 (58.3)	1044 (51.5)	
Men	997 (48.4)	15 (41.7)	982 (48.5)	
Education, n (%)				>.99
No high school	229 (11.1)	4 (11.1)	225 (11.1)	
High school	386 (18.7)	7 (19.4)	379 (18.7)	
Some college	551 (26.7)	9 (25.0)	542 (26.8)	
College and higher	896 (43.4)	16 (44.4)	880 (43.4)	
Neuropsychological test score, mean (SD)				
Logical Memory—Immediate Recall	12 (3)	10 (3)	13 (3)	<.001
Logical Memory—Delayed Recall	12 (4)	8 (3)	12 (4)	<.001
Logical Memory—Recognition	10 (1)	9 (2)	10 (1)	.002
Visual Reproduction—Immediate Recall	8 (3)	5 (2)	8 (3)	<.001
Visual Reproduction—Delayed Recall	8 (3)	3 (2)	8 (3)	<.001
Visual Reproduction—Recognition	3 (1)	2 (1)	3 (1)	<.001
Paired Associate Learning—Immediate Recall	15 (4)	11 (3)	15 (3)	<.001
Paired Associate Learning—Delayed Recall	9 (1)	7 (2)	9 (1)	<.001
Paired Associate Learning—Recognition	10 (2)	9 (2)	10 (2)	<.001
Digit Span—Forward	7 (1)	6 (1)	7 (1)	<.001
Digit Span—Backward	5 (1)	4 (1)	5 (1)	<.001
Similarities	17 (3)	14 (4)	17 (3)	<.001
Boston Naming Test—30-item version	26 (7)	23 (6)	26 (7)	<.001
Trail Making Test A (seconds)	32 (17)	47 (13)	32 (17)	<.001
Trail Making Test B (seconds)	88 (73)	213 (131)	86 (70)	<.001
Hooper Visual Organization Test	25 (3)	22 (3)	26 (3)	<.001
Verbal Fluency	41 (12)	31 (13)	41 (12)	<.001
Verbal Fluency—Animal	19 (6)	13 (5)	19 (6)	<.001

A total of 105 distinct dCDT features were derived from each dCDT drawing. Associations between each individual dCDT feature and 18 neuropsychological tests assessing different cognitive functions are shown in Table S1 (Multimedia Appendix 1). In addition, dCDT features that were significantly associated different cognitive domains were summarized in Table S2 (Multimedia Appendix 1). On average, each neuropsychological test was associated with 50 dCDT features after adjusting for multiple testing.

The weighted composite scores built from significant dCDT features for each neuropsychological test were all significantly associated with their corresponding neuropsychological tests (Table 2).

Table 2. Association between digital Clock Drawing Test (dCDT) composite scores and neuropsychological tests.

Neuropsychological test	Significant dCDT features, n	Participants, n	Effect size	Standard error	Bonferroni-corrected *P* value[a]
Logical Memory—Immediate Recall	48	2048	0.0625	0.0061	7.4×10^{-24}
Logical Memory—Delayed Recall	54	2047	0.0571	0.0059	8.1×10^{-22}
Logical Memory—Recognition	28	2037	0.0681	0.0086	4.2×10^{-15}
Visual Reproduction—Immediate Recall	62	2049	0.0592	0.0041	3.0×10^{-45}
Visual Reproduction—Delayed Recall	61	2048	0.0596	0.0042	8.0×10^{-44}
Visual Reproduction—Recognition	69	2043	0.0548	0.0044	5.3×10^{-35}
Paired Associate Learning—Immediate Recall	32	1991	0.0966	0.0104	5.8×10^{-20}
Paired Associate Learning—Delayed Recall	50	2026	0.0612	0.0064	5.2×10^{-21}
Paired Associate Learning—Recognition	13	2062	0.1561	0.0287	5.8×10^{-8}
Digit Span—Forward	32	2053	0.0773	0.0092	8.4×10^{-17}
Digit Span—Backward	35	2032	0.0780	0.0090	8.0×10^{-18}
Trail Making Test A	81	2043	0.0404	0.0019	6.5×10^{-87}
Trail Making Test B	86	1999	0.0428	0.0026	1.0×10^{-56}
Similarities	46	2034	0.0783	0.0072	4.1×10^{-27}
Hooper Visual Organization Test	66	2007	0.0583	0.0041	2.5×10^{-44}
Boston Naming Test—30-item version	47	2062	0.0615	0.0063	2.7×10^{-22}
Verbal Fluency	66	2007	0.0523	0.0048	4.0×10^{-27}
Verbal Fluency—Animal	31	2062	0.0637	0.0068	1.2×10^{-20}

[a] All *P* values remained significant after Bonferroni correction ($P<2.8\times10^{-3}$).

Eight neuropsychological tests were significantly associated with mild cognitive impairment ($P<2.8\times10^{-3}$), including Visual Reproduction—Delayed Recall, Visual Reproduction—Immediate Recall, Visual Reproduction—Recognition, Paired Associate Learning—Immediate Recall, Paired Associate Learning—Delayed Recall, Digit Span—Backward, Trail Making Test B, and Logical Memory—Delayed Recall (Table 3). All dCDT composite scores were significantly associated with mild cognitive impairment ($P<2.8\text{x}10^{-3}$).

Table 3. Association between neuropsychological tests and digital Clock Drawing Test (dCDT) composite scores and mild cognitive impairment.

Test type	Neuropsychological tests			dCDT composite scores		
	Coefficient estimate	Standard error	*P* value[a]	Coefficient estimate	Standard error	*P* value
Logical Memory—Immediate Recall	–0.1256	0.0459	6.2×10^{-3}	–0.0590	0.0138	2.0×10^{-5}
Logical Memory—Delayed Recall	–0.1401	0.0437	1.4×10^{-3}	–0.0523	0.0124	2.6×10^{-5}
Logical Memory—Recognition	–0.2469	0.1067	2.1×10^{-2}	–0.2131	0.0504	2.3×10^{-5}
Visual Reproduction—Immediate Recall	–0.3044	0.0712	1.9×10^{-5}	–0.0498	0.0116	1.8×10^{-5}
Visual Reproduction—Delayed Recall	–0.3558	0.0757	2.6×10^{-6}	–0.0465	0.0110	2.2×10^{-5}
Visual Reproduction—Recognition	–0.4896	0.1519	1.3×10^{-3}	–0.1357	0.0333	4.5×10^{-5}
Paired Associate Learning—Immediate Recall	–0.1939	0.0542	3.5×10^{-4}	–0.1157	0.0236	9.2×10^{-7}
Paired Associate Learning—Delayed Recall	–0.3861	0.1100	4.5×10^{-4}	–0.1575	0.0349	6.4×10^{-6}
Paired Associate Learning—Recognition	–0.0860	0.0675	2.0×10^{-1}	–0.5349	0.1334	6.0×10^{-5}
Digit Span—Forward	–0.2922	0.1549	5.9×10^{-2}	–0.2417	0.0560	1.6×10^{-5}
Digit Span—Backward	–0.6303	0.1858	6.9×10^{-4}	–0.2493	0.0551	6.1×10^{-6}
Trail Making Test A	0.0022	0.0050	6.6×10^{-1}	0.0033	0.0010	1.0×10^{-3}
Trail Making Test B	0.0039	0.0012	1.1×10^{-3}	0.0010	0.0003	3.3×10^{-4}
Similarities	–0.1118	0.0421	7.9×10^{-3}	–0.0668	0.0154	1.4×10^{-5}
Hooper Visual Organization Test	–0.1030	0.0435	1.8×10^{-2}	–0.0449	0.0102	1.2×10^{-5}
Boston Naming Test—30-item version	–0.0236	0.0212	2.7×10^{-1}	–0.0300	0.0070	1.8×10^{-5}
Verbal Fluency	–0.0396	0.0161	1.4×10^{-2}	–0.0115	0.0029	6.3×10^{-5}
Verbal Fluency—Animal	–0.0727	0.0283	1.0×10^{-2}	–0.0366	0.0088	3.1×10^{-5}

Discussion

Neuropsychological tests have been widely used in the assessment of cognitive performance. All 18 neuropsychological tests, for the assessment of multiple cognitive domains, were significantly associated with an average of 50 dCDT features (range 13 to 86 features), and dCDT composite scores were significantly associated with mild cognitive impairment compared to normal cognition.

The CDT examines a wide range of cognitive abilities [5]. The command condition requires that an individual first understand the verbal command, recall all clock related attributes from semantic memory, understand the visuospatial relationships between clock features, and execute the command using necessary mental planning and visuoconstructional abilities. For the copy condition, successful performance requires that an individual recognize the visuospatial attributes in the model to be copied and then marshal the necessary executive abilities to execute output in an organized fashion. However, the standard pencil-and-paper CDT for dementia assessment is usually subjective and time intensive. Given only a limited number of features can be scored, the standard CDT has relatively inferior sensitivity and variable specificity for mild or questionable dementia [2]. On the other hand, the dCDT provides comprehensive and objective assessment of multiple cognitive domains with far greater time efficiency [29,30]. It was reported that the total completion time of dCDT was positively correlated with cognitive functions, whereas the post–clock face latency and pre–first-hand latency were negatively associated with the working memory and processing speed [31]. Our study extended this work by including more than 100 dCDT features and assessing their association with 18 different neuropsychological tests. The composite scores built from dCDT features were significantly associated (all *P*<.001) with multiple neuropsychological tests, such as Trail Making Test A, Trail Making Test B, Hooper Visual Organization Test, and Visual Reproduction subtests. Our results suggested that dCDT composite scores represent better surrogates for their corresponding neuropsychological tests than individual dCDT features. The results also underscore the psychometric characteristics of the dCDT for measuring multiple cognitive domains, such as attention, executive function, visuoperceptual organization, and visual memory, findings which are consistent with those of prior studies [1,5].

A complete neuropsychological test protocol must be administered by a trained rater and interpreted by a neuropsychological expert, which takes at least 45 minutes; therefore, financial and medical resource requirements limit the application of a complete neuropsychological test protocol in general clinics. In contrast, the dCDT is much more convenient, and the test generally takes less than 2 minutes. It is also worth noting that some patients with mild cognitive impairment may show normal performance in some of cognitive domains of neuropsychological test, demonstrating reduced sensitivity of these neuropsychological tests in detecting mild cognitive impairment for some patient groups [32]. As dCDT composite scores were derived from a combination of multiple features associated with neuropsychological tests, they have the potential to identify more subtle cognitive impairment than individual neuropsychological tests. In an earlier study, Dion et al [31] analyzed 202 older adults without dementia and found that participants with mild cognitive impairment tended to take more time to complete the entire test—more “Think than Ink” (ie, percentage of time thinking vs percentage of time drawing)—and drew smaller clock face areas than those drawn by participants with normal cognition. In another study consisting of 138 patients with mild cognitive impairment and amnesia, 106 patients with mild Alzheimer disease, and 137 normal cognition participants; a tablet-based dCDT provided a slightly higher diagnostic accuracy for patients with mild cognitive impairment and amnesia than the CERAD (Consortium to Establish a Registry for Alzheimer's Disease) total score (81.5% vs 77.5%) [33]. The dCDT features have also been used to differentiate between other neurological diseases, such as memory impairment disorders, vascular cognitive disorders, and Parkinson disease [17].

Several study limitations merit consideration. First, this was a cross-sectional study, which cannot reveal temporal relationship between dCDT performance and mild cognitive impairment. It would be interesting to perform longitudinal analysis to investigate early cognitive markers of dCDT features that predict future cognitive decline. Second, only a moderate number of patients with mild cognitive impairment were included. The number of patients with dementia was even smaller, and therefore, patients with dementia were excluded. Third, neuropsychological tests were used to diagnose mild cognitive impairment, which possibly caused some circularity and overestimated diagnostic performance of neuropsychological tests. On the other hand, neither dCDT features nor derived composite scores were used for the mild cognitive impairment diagnosis, which reduced the bias of potential overestimation. Fourth, due to the increasing exposure to digitalized clock displays, a recent study found that some participants drew digital clocks instead of analog clocks required by the test [34]. It is thus important to continue to explore novel cognitive assessment strategies to better capture new features from different neuropsychological tests to avoid potential bias caused by this new technology trend. Finally, yet importantly, FHS participants were mostly of European ancestry and English speakers, therefore, the applicability of these findings to populations of other race and ethnicity is unknown. Notwithstanding these limitations, our study had several strengths. We studied the association between dCDT and a standard epidemiologic neuropsychological test protocol with community-based FHS study data. FHS data have been collected consistently with rigorous quality control and clinical diagnosis by consensus review. Notably, unlike tablet-based apps, the digital pen used in our study offers an almost identical user experience as that of a traditional ballpoint pen; no extra training is needed, which is particularly important for older adult participants who might be unfamiliar with new digital technologies. The performance was thus less likely distorted [10,11].

Associations between dCDT features and standard neuropsychological test data, as well as composite scores from dCDT features as an alternative to neuropsychological tests for the classification of mild cognitive impairment, from more than 2000 participants from a large community-based cohort suggest the potential of dCDT as a cost-effective and easy-to-administer tool for general practitioners, with potential for use in low-resource countries or regions where clinical dementia expertise is limited.

Acknowledgments

This work was supported by the National Heart, Lung, and Blood Institute (Framingham Heart Study’s contract N01-HC-25195; HHSN268201500001I), the National Institute on Aging (AG008122, AG016495, AG033040, AG054156, AG049810, AG062109, AG068753), and Pfizer. This work was also supported by the Alzheimer’s Association (AARG-NTF-20-643020) and American Heart Association (20SFRN35360180). Support for JY was provided by the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2016I2M1004).

We thank the Framingham Heart Study participants for their decades of dedication and the staff for their hard work in collecting and preparing the data.

Authors' Contributions

RA and DJL designed the study protocol. HL, JY, and RA designed the analysis strategy, and HL performed the statistical analyses. JY drafted the manuscript. CK and AFAA prepared the data. RA, SD and SA performed dementia review. All authors critically reviewed and approved the final manuscript.

Conflicts of Interest

RA is a scientific advisor to Signant Health and consultant to Biogen.

Multimedia Appendix 1
Supplementary tables.
[PDF File (Adobe PDF File), 292 KB - jmir_v23i6e27407_app1.pdf]

References

1. Agrell B, Dehlin O. The clock-drawing test. Age Ageing 1998;27(3):399. [doi: 10.1093/ageing/afs149] [Medline: 23144287]
2. Pinto E, Peters R. Literature review of the clock drawing test as a tool for cognitive screening. Dement Geriatr Cogn Disord 2009 Feb 19;27(3):201-213. [doi: 10.1159/000203344] [Medline: 19225234]
3. Shulman KI. Clock-drawing: is it the ideal cognitive screening test? Int J Geriat Psychiatry 2000 Jun;15(6):548-561. [doi: 10.1002/1099-1166(200006)15:6<548::aid-gps242>3.0.co;2-u] [Medline: 10861923]
4. Freedman M, Leach L, Kaplan E, Shulman K, Delis D. Clock Drawing: A Neuropsychological Analysis. United States: Oxford University Press; 1994.
5. Libon DJ, Swenson RA, Barnoski EJ, Sands LP. Clock drawing as an assessment tool for dementia. Arch Clin Neuropsychol 1993 Jan 01;8(5):405-415. [doi: 10.1093/arclin/8.5.405] [Medline: 14589710]
6. Cosentino S, Jefferson A, Chute DL, Kaplan E, Libon DJ. Clock drawing errors in dementia: neuropsychological and neuroanatomical considerations. Cogn Behav Neurol 2004 Jun;17(2):74-84. [doi: 10.1097/01.wnn.0000119564.08162.46] [Medline: 15453515]
7. Kaplan E. The process approach to neuropsychological assessment of psychiatric patients. J Neuropsychiatry Clin Neurosci 1990;2(1):72-87. [doi: 10.1176/jnp.2.1.72] [Medline: 2136065]
8. Kaplan E. A process approach to neuropsychological assessment. In: Boll T, Bryant BK, editors. Clinical Neuropsychology and Brain Function: Research, Measurement, and Practice. Washington, DC: American Psychological Association; 1988:127-167.
9. Penney D, Davis R, Libon D, Lamar M, Price C, Swenson R. The Digital Clock Drawing Test (dCDT)-II: a new computerized quantitative system. Boston, MA; 2011 Presented at: 40th Annual Meeting of the International Neuropsychological Society; February 15-18; Montreal, Canada URL: http://rationale.csail.mit.edu/clocksketch/Publications/INS2011-dCDTII.pdf
10. Davis R, Penney D, Pittman D, Libon D, Swenson R, Kaplan E. The Digital Clock Drawing Test (dCDT) I: development of a new computerized quantitative system. 2010 Presented at: 39th Annual Meeting of the International Neuropsychological Society; February 2-5; Boston, MA URL: http://rationale.csail.mit.edu/clocksketch/Publications/INS2011-dCDTI.pdf
11. Davis R, Libon D, Au R, Pitman D, Penney D. THink: Inferring Cognitive Status from Subtle Behaviors. AI Mag 2015 Sep 28;36(3):49-60. [doi: 10.1609/aimag.v36i3.2602]
12. Cohen J, Penney DL, Davis R, Libon DJ, Swenson RA, Ajilore O, et al. Digital Clock Drawing: differentiating "thinking" versus "doing" in younger and older adults with depression. J Int Neuropsychol Soc 2014 Oct;20(9):920-928 [FREE Full text] [doi: 10.1017/S1355617714000757] [Medline: 25222513]
13. Piers RJ, Devlin KN, Ning B, Liu Y, Wasserman B, Massaro JM, et al. Age and graphomotor decision making assessed with the digital clock drawing test: the Framingham Heart Study. J Alzheimers Dis 2017;60(4):1611-1620 [FREE Full text] [doi: 10.3233/JAD-170444] [Medline: 29036819]
14. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 2014 Mar 15;383(9921):999-1008 [FREE Full text] [doi: 10.1016/S0140-6736(13)61752-3] [Medline: 24084292]
15. Tsao CW, Vasan RS. Cohort profile: the Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol 2015 Dec;44(6):1800-1813. [doi: 10.1093/ije/dyv337] [Medline: 26705418]
16. Wolf PA. Contributions of the Framingham Heart Study to stroke and dementia epidemiologic research at 60 years. Arch Neurol 2012 May;69(5):567-571 [FREE Full text] [doi: 10.1001/archneurol.2011.977] [Medline: 22213410]
17. Souillard-Mandar W, Davis R, Rudin C, Au R, Libon DJ, Swenson R, et al. Learning classification models of cognitive conditions from subtle behaviors in the digital clock drawing test. Mach Learn 2016 Mar;102(3):393-441 [FREE Full text] [doi: 10.1007/s10994-015-5529-5] [Medline: 27057085]
18. Au R, Seshadri S, Wolf PA, Elias MF, Elias PK, Sullivan L, et al. New norms for a new generation: cognitive performance in the Framingham offspring cohort. Exp Aging Res 2004 Oct;30(4):333-358. [doi: 10.1080/03610730490484380] [Medline: 15371099]
19. Wechsler D, Stone CP. Wechsler Memory Scale. New York, NY: The Psychological Corporation; 1948.
20. Wechsler D. Wechsler Adult Intelligence Scale (WAIS). New York: The Psychological Corporation; 1955.
21. Kaplan E, Goodglass H, Weintraub S, Segal O. Boston Naming Test. Philadephia: Lea & Febiger; 1983.
22. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 1958 Dec;8(7):271-276. [doi: 10.2466/pms.8.7.271-276]
23. Hooper HE. The Hooper Visual Organization Test: Manual. Los Angeles: Western Psychological Service; 1958.
24. Benton A, Hamsher K, Sivan A. Multilingual Aphasia Examination. Iowa City: University of Iowa; 1976.
25. Goodglass H, Kaplan E. The Assessment of Aphasia and Related Disorders 2nd Edition. Philadelphia: Lea & Febiger; 1983.

26. Seshadri S, Wolf PA, Beiser A, Au R, McNulty K, White R, et al. Lifetime risk of dementia and Alzheimer's disease. the impact of mortality on risk estimates in the Framingham Study. Neurology 1997 Dec;49(6):1498-1504. [doi: 10.1212/wnl.49.6.1498] [Medline: 9409336]
27. Au R, Piers RJ, Devine S. How technology is reshaping cognitive assessment: lessons from the Framingham Heart Study. Neuropsychology 2017 Nov;31(8):846-861 [FREE Full text] [doi: 10.1037/neu0000411] [Medline: 29376667]
28. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982 Jun;140:566-572. [Medline: 7104545]
29. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. 2014. URL: https://www.apa.org/science/programs/testing/standards [accessed 2021-05-27]
30. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Clin Neuropsychol 2012;26(2):177-196 [FREE Full text] [doi: 10.1080/13854046.2012.663001] [Medline: 22394228]
31. Dion C, Arias F, Amini S, Davis R, Penney D, Libon DJ, et al. Cognitive correlates of digital clock drawing metrics in older adults with and without mild cognitive impairment. J Alzheimers Dis 2020;75(1):73-83 [FREE Full text] [doi: 10.3233/JAD-191089] [Medline: 32250300]
32. Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004 Sep;256(3):183-194 [FREE Full text] [doi: 10.1111/j.1365-2796.2004.01388.x] [Medline: 15324362]
33. Müller S, Herde L, Preische O, Zeller A, Heymann P, Robens S, et al. Diagnostic value of digital clock drawing test in comparison with CERAD neuropsychological battery total score for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. Sci Rep 2019 Mar 05;9(1):3543 [FREE Full text] [doi: 10.1038/s41598-019-40010-0] [Medline: 30837580]
34. Ilardi CR, Garofalo E, Chieffi S, Gamboz N, La Marra M, Iavarone A. Daily exposure to digital displays may affect the clock-drawing test: from psychometrics to serendipity. Neurol Sci 2020 Dec;41(12):3683-3690. [doi: 10.1007/s10072-020-04498-z] [Medline: 32506359]

Abbreviations

CDT: Clock Drawing Test
dCDT: digital Clock Drawing Test
FHS: Framingham Heart Study

Edited by G Eysenbach; submitted 24.01.21; peer-reviewed by K Jekel, M Nitsch, J Heo; comments to author 12.02.21; revised version received 05.03.21; accepted 27.04.21; published 08.06.21.

Please cite as:
Yuan J, Libon DJ, Karjadi C, Ang AFA, Devine S, Auerbach SH, Au R, Lin H
Association Between the Digital Clock Drawing Test and Neuropsychological Test Performance: Large Community-Based Prospective Cohort (Framingham Heart Study)
J Med Internet Res 2021;23(6):e27407
URL: https://www.jmir.org/2021/6/e27407
doi:10.2196/27407
PMID:34100766

Original Paper

Acceptability of the Medication Event Reminder Monitor for Promoting Adherence to Multidrug-Resistant Tuberculosis Therapy in Two Indian Cities: Qualitative Study of Patients and Health Care Providers

Beena E Thomas[1*], PhD, MSW; J Vignesh Kumar[1], PhD, MSW; Murugesan Periyasamy[1], MA, PGDC; Amit Subhash Khandewale[1], MSW; J Hephzibah Mercy[1], MSW; E Michael Raj[1], MSW; S Kokila[1], MSc; Apurva Shashikant Walgude[1], MA; Gunjan Rahul Gaurkhede[1], MA, MSW; Jagannath Dattatraya Kumbhar[1], MSW; Senthanro Ovung[1], MSW; Mariyamma Paul[1], MSW; B Sathyan Rajkumar[1], MA; Ramnath Subbaraman[2,3,4*], MSc, MD, FACP

[1] Department of Social and Behavioural Research, ICMR-National Institute for Research in Tuberculosis, Chennai, India
[2] Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, United States
[3] Center for Global Public Health, Tufts University School of Medicine, Boston, MA, United States
[4] Division of Geographic Medicine and Infectious Diseases, Tufts Medical Center, Boston, MA, United States
[*] these authors contributed equally

Corresponding Author:
Ramnath Subbaraman, MSc, MD, FACP
Department of Public Health and Community Medicine
Tufts University School of Medicine
136 Harrison Ave
MV237
Boston, MA, 02111
United States
Phone: 1 6176360985
Email: ramnath.subbaraman@tufts.edu

Abstract

Background: Patients with multidrug-resistant tuberculosis (MDR-TB) face challenges adhering to medications, given that treatment is prolonged and has a high rate of adverse effects. The Medication Event Reminder Monitor (MERM) is a digital pillbox that provides pill-taking reminders and facilitates the remote monitoring of medication adherence.

Objective: This study aims to assess the MERM's acceptability to patients and health care providers (HCPs) during pilot implementation in India's public sector MDR-TB program.

Methods: From October 2017 to September 2018, we conducted qualitative interviews with patients who were undergoing MDR-TB therapy and were being monitored with the MERM and HCPs in the government program in Chennai and Mumbai. Interview transcripts were independently coded by 2 researchers and analyzed to identify the emergent themes. We organized findings by using the Unified Theory of Acceptance and Use of Technology (UTAUT), which outlines 4 constructs that predict technology acceptance—performance expectancy, effort expectancy, social influence, and facilitating conditions.

Results: We interviewed 65 patients with MDR-TB and 10 HCPs. In patient interviews, greater acceptance of the MERM was related to perceptions that the audible and visual reminders improved medication adherence and that remote monitoring reduced the frequency of clinic visits (performance expectancy), that the device's organization and labeling of medications made it easier to take them correctly (effort expectancy), that the device facilitated positive family involvement in the patient's care (social influences), and that remote monitoring made patients feel more *cared for* by the health system (facilitating conditions). Lower patient acceptance was related to problems with the durability of the MERM's cardboard construction and difficulties with portability and storage because of its large size (effort expectancy), concerns regarding stigma and the disclosure of patients' MDR-TB diagnoses (social influences), and the incorrect understanding of the MERM because of suboptimal counseling (facilitating conditions). In their interviews, HCPs reported that MERM implementation resulted in fewer in-person interactions with patients and thus allowed HCPs to dedicate more time to other tasks, which improved job satisfaction.

Conclusions: Several features of the MERM support its acceptability among patients with MDR-TB and HCPs, and some barriers to patient use could be addressed by improving the design of the device. However, some barriers, such as disease-related stigma, are more difficult to modify and may limit use of the MERM among some patients with MDR-TB. Further research is needed to assess the accuracy of MERM for measuring adherence, its effectiveness for improving treatment outcomes, and patients' sustained use of the device in larger scale implementation.

(J Med Internet Res 2021;23(6):e23294) doi:10.2196/23294

KEYWORDS

tuberculosis; drug-resistant; medication adherence; mHealth; digital adherence technologies; India

Introduction

Background

Multidrug-resistant tuberculosis (MDR-TB) is a major global challenge to tuberculosis (TB) control. In 2018, approximately 484,000 people worldwide were estimated to have developed MDR-TB, including approximately 130,000 people in India [1]. Despite considerable advances in therapy in the last decade, treatment outcomes remain poor for individuals with MDR-TB, with treatment success rates of 56% worldwide and 48% in India for the 2017 patient cohort [1]. Although some of the variability in treatment outcomes may be attributable to the composition of the patient's drug regimen [2], suboptimal adherence to medications may be another critical problem contributing to poor MDR-TB treatment outcomes.

Successful adherence for diseases with a prolonged treatment course, such as MDR-TB, requires a high level of dosing implementation (ie, taking a medication dose on a given day) and persistence (ie, taking medications for the entire duration of therapy [3]). Factors contributing to nonadherence are complex and include therapy-related (eg, toxicities [4]), psychosocial (eg, alcohol use [5], depression [6], and stigma [7]), structural (eg, distance from clinics and medication costs [8,9]), and health system–related challenges (eg, poor user experience with the health system). Patients with MDR-TB face particularly high levels of drug toxicity [4] and psychosocial barriers, including depression, substance use disorders, stigma, and discrimination [10]. These challenges may lead to poor outcomes and increased transmission of drug-resistant strains.

As such, there is an urgent need for new strategies to improve medication adherence in patients with MDR-TB. Many TB programs have historically used directly observed therapy (DOT) to monitor adherence; however, recent studies have questioned the efficacy of this strategy for improving clinical outcomes [11-13] and raised concerns that DOT adversely affects patient autonomy [14,15]. Limited autonomy with DOT may be greater for patients receiving MDR-TB treatment, given the prolonged course of therapy required. In addition, recent recommendations favoring the use of regimens containing only oral medications may decrease the required frequency of clinic visits for patients with MDR-TB if they are allowed to take therapy without in-person observation [16].

In recent years, driven by the global expansion of cellular networks, there has been a growing use of digital adherence technologies (DATs) as alternative approaches for monitoring adherence to TB medications [3]. These technologies, which include cellphones, digital pillboxes, and ingestion sensors, have the potential to improve clinical outcomes via multiple pathways [3]. Although there are numerous DATs aimed at addressing nonadherence in patients with drug-susceptible TB [3], few have attempted to address the more complex medication regimens taken by patients with MDR-TB [17].

The Medication Event Reminder Monitor (MERM) is a digital pillbox that has been designed to monitor MDR-TB treatment in resource-constrained settings by using relatively affordable evriMED technology produced by Wisepill Technologies. This system is designed to be used with multiple blister-packaged medications in MDR-TB regimens, incorporates visual and audible reminders for daily dosing and refills, compiles dosing histories by capturing data on pillbox opening as a proxy for dose ingestion, and transmits these data to a server so that health care providers (HCPs) can remotely visualize patients' dosing histories. By providing near real-time adherence data, the MERM may facilitate the identification of high-risk patients and prompt early intervention by HCPs to reduce nonadherence. When compared with facility-based DOT, in which patients travel to clinics to be observed taking medications, monitoring by using the MERM may reduce the required frequency of patient visits to TB clinics.

Pilot studies of digital pillboxes conducted in Uganda and China with patients with drug-susceptible TB and in South Africa with patients with MDR-TB have shown these devices to have relatively high acceptability [17-19]. A cluster randomized trial conducted in China with patients with drug-susceptible TB found digital pillboxes to be effective in reducing the percentage of patient months with high nonadherence [20]. However, subsequent studies on the large-scale implementation of these digital pillboxes in China have revealed challenges in their uptake. For example, after accounting for patients who were not eligible to use these pillboxes, refused to use them, withdrew from using them early in treatment, or were shifted to monitoring with DOT, only approximately 49% and 39% of patients with TB used digital pillboxes in a sustained manner in a single province [21] and 30 counties [22], respectively, in China. In addition, a study of the MERM conducted in Vietnam with patients with drug-susceptible TB found that only approximately half of the patients used the device as intended, with many separating the time when the pillbox was opened from the time that doses were ingested, because of concerns about the device's portability [23]. These existing studies evaluating the use of digital pillboxes reveal variability in patient acceptance in different contexts and highlight a relative paucity of data on the use of these devices for patients with MDR-TB.

Objectives

In this study, based on qualitative interviews with patients and HCPs, we evaluated the MERM's ability to monitor adherence to MDR-TB therapy during pilot rollout in the government's National TB Elimination Program (NTEP) in 2 Indian cities. Although this novel strategy has potential advantages, previous research has not been conducted in India to evaluate whether patients will accept and use the MERM, to identify potential modifiable and nonmodifiable barriers to its acceptability, and to understand how its implementation might impact HCP work efficiency and quality of care. Understanding the acceptability of MERM during pilot implementation is also important because recent studies of other DATs in India suggest that suboptimal acceptability and use by patients could reduce the accuracy of these technologies for measuring adherence [24], which might in turn greatly reduce the value of real-time adherence data for guiding interventions. We analyze our findings using the Unified Theory of Acceptance and Use of Technology (UTAUT), a framework that synthesizes constructs that predict engagement with novel technologies [25,26].

Methods

Ethical Approvals

This protocol received approval from the Indian Council of Medical Research-National Institute for Research in TB Institutional Ethics Committee (FWA00005104) on March 7, 2017. The study was approved by the Brigham and Women's Hospital Institutional Review Board (FWA00000484) on January 31, 2017, and the Tufts Health Sciences Institutional Review Board (FWA00004517) on June 6, 2018. Written informed consent was obtained from all the participants.

Study Setting

This study was conducted in 2 Indian cities with a high TB burden in the general population [27,28]: Chennai (estimated population of 7.1 million in metropolitan area), in the southern Indian state of Tamil Nadu, and Mumbai (estimated population of 18.4 million in the metropolitan area), in the western Indian state of Maharashtra. Mumbai in particular has one of the world's most severe urban epidemics of drug-resistant TB [29-31]. All patients in Chennai spoke Tamil, and all patients in Mumbai spoke Hindi or Marathi. Consistent with the broader population of patients with TB who seek government care in India, more than half of whom earn less than US $2 a day of income [32], patients with MDR-TB in these 2 cities come from socioeconomically disadvantaged backgrounds.

MERM Implementation

Medications are dispensed in blister packs, and each drug is placed in a different partitioned compartment within the MERM, which facilitates storage and organization of the multiple medications that comprise MDR-TB regimens (Figure 1). In India's pilot implementation, the container and internal partitions were made of cardboard. The device was provided to patients at different time points in the continuation phase of therapy when injectable agents had generally been discontinued and patients were only taking oral medications. At the time of our study, most patients with MDR-TB in India's NTEP were placed on a standardized drug regimen for a treatment course lasting 24 to 27 months, with the continuation phase consisting of levofloxacin, ethionamide, cycloserine, and ethambutol taken once daily [33]. Patients who were provided the MERM subsequently used it for the remaining duration of treatment.

The MERM was programmed to provide audible and visual reminders to take medications at a specific time each day, per patient and HCP preference. The visual reminder consisted of a blinking green light corresponding to a label encouraging the patient to take a dose; separate yellow and red lights blink to alert patients about the need to refill medications and replace the MERM's battery, respectively. The audio reminder consisted of a ringing sound that occurred at the same time as the visual dose-taking reminder.

The device contained a removable battery-powered module. Triggered by a magnetic sensor, this module captures and stores data each time the container is opened as a proxy for medication ingestion. These data on patient engagement with the MERM were transmitted every 72 hours using cellular networks and recorded on a computer server. HCPs could log into an app on a mobile device or a website, where each patient's adherence history was presented as a color-coded calendar in which green suggested that the MERM was opened on a given day (suggesting dose ingestion), whereas red suggested that the device was not opened (suggesting that the dose was not ingested).

These dosing histories were meant to help HCPs have individualized discussions with patients regarding their adherence. In addition, a series of probable missed doses would result in automated SMS text messaging notifications prompting HCPs to intervene with these patients, who might be at a higher risk for unfavorable outcomes.

Figure 1. The Medication Event and Reminder Monitor in a cardboard version used for the initial rollout among patients with multidrug-resistant tuberculosis in India. The device includes partitions for organizing medications, medication labels inside the box lid, and a digital module that provides reminders and captures adherence data. This cardboard version was provided by Wisepill Technologies.

Recruitment of Study Participants and Collection of Qualitative Data

Interviews were conducted by 3 field researchers in Mumbai (1 man and 2 women) and 3 field researchers in Chennai (2 men and 1 woman), all of whom had a master's degree in social work or another social science field who underwent a 2-day training in qualitative interviews at the National Institute for Research in TB in Chennai. Study participants included patients with MDR-TB and HCPs. We use the term *MDR-TB* to describe patients with confirmed resistance to isoniazid and rifampin as well as individuals who were diagnosed as having rifampin-resistant TB using Xpert MTB/RIF (Cepheid, Inc) because patients with rifampin-resistant TB in India are treated as having likely MDR-TB. Data collection was conducted a few months after the MERM was introduced in Mumbai and Chennai for monitoring patients with MDR-TB from October 2017 to September 2018. Before the rollout of the MERM, HCPs were extensively trained on the appropriate use of the MERM. HCPs dispensed medication refills in the MERM for patients with MDR-TB in the continuation phase of therapy after any injectable agents (eg, kanamycin) were discontinued.

Field researchers met patients at MDR-TB clinics, where patients were screened for inclusion in the study. At these clinics, HCPs were recruited for the study, including health visitors (individuals with at least a high school level of education who monitor therapy), senior treatment supervisors (individuals with at least a high school level of education who supervise health visitors), medical officers (doctors with an MBBS or higher degree), and district TB officers (doctors who supervise TB care across a district).

For patients, an unannounced home visit was made at least 3 weeks after enrollment into the study to conduct the qualitative in-depth interviews regarding the MERM, which lasted about 45 minutes. A pill count was also conducted to better understand the patients' adherence to medications. We ensured a minimum of 3 weeks lapsed between when a patient was consented to the study and when the unannounced home visit was conducted. This time gap minimized the impact of temporary changes in medication adherence that may result from the patient knowing that he or she will be visited as part of the study (ie, the Hawthorne effect). For HCPs, interviews lasted about 30-45 minutes and were conducted in a private space in the TB clinic.

To ensure uniformity in data collection, separate patient and HCP interview guides, each consisting of open-ended and semistructured questions with follow-up probes, were used to conduct the interviews. Examples of questions from the patient interview guide are provided in Multimedia Appendix 1. The interview questions had the goal of assessing key constructs in the UTAUT framework. Interviews were conducted in Tamil, Hindi, or Marathi and audio recorded. They were later transcribed and translated to produce deidentified English language transcripts. To ensure translation accuracy, one-fourth of the English transcripts were evaluated against the original audio recordings for correctness and completeness.

Analytical Framework: UTAUT

The UTAUT integrates constructs from previous literature on technology acceptance into a single framework [25,26]. Of these constructs, 3—performance expectancy, effort expectancy, and social influences—help to predict the intention of individuals to use a technology, which is necessary, but not sufficient, to result in actual use. Performance expectancy refers to the perceived usefulness of technology for users. For example, for patients with MDR-TB, this may refer to the extent to which the MERM is perceived to improve their medical care, whereas for HCPs, it may refer to the extent to which it is perceived to improve the quality or efficiency of their work. Effort expectancy refers to the ease of using the technology. For patients, this may refer to the effort required to correctly understand the different functions of the digital pillbox (eg, storage function and reminders), whereas for HCPs, this may refer to the effort required to use and understand the web-based adherence dashboard and SMS text messaging reminders that notify HCPs about nonadherent patients. Social influences refer to how other individuals may influence a person's acceptance and use of technology. For patients, this may include family members or community residents, whereas for HCPs, this may include coworkers. Of these 3 constructs, performance expectancy has the strongest influence on the intention to use a technology [25].

Facilitating conditions, the fourth construct in the UTAUT, is thought to directly affect the actual use of a new technology. Facilitating conditions comprise the underlying infrastructure to enable the use of a technology. For patients with MDR-TB, we interpreted this broadly to include factors in the TB program, such as the quality of counseling regarding the MERM provided to patients and any outreach to patients by HCPs that might have been prompted by adherence data from the MERM. For HCPs, we interpret this to include the quality of training they received before the rollout of the MERM and any higher-level support they received during the rollout process.

Analysis of Qualitative Data

Transcribed interviews were coded using a thematic approach and analyzed using Dedoose software (version 8.0.35; SocioCultural Research Consultants, LLC). The study team first identified possible codes (ie, themes) related to the central research question from the data collected by using the qualitative interview guides as a foundation and the UTAUT as an organizing framework. The transcripts were then independently coded by 2 researchers for relevant themes using descriptive content analysis. In parallel, the researchers tracked new codes that were added to the coding scheme to describe unexpected themes that emerged. The two researchers frequently met to reconcile inconsistencies in the application of codes and to ensure that emergent codes were added to the coding scheme. Because all coding differences were reconciled by consensus, we did not assess the interrater reliability between the coders.

We analyzed the data to identify emergent themes that could influence the acceptability and use of MERM. Emergent themes were organized within the 4 UTAUT constructs, and illustrative quotations were selected for each theme. In reporting our findings, we follow the principles of qualitative research by avoiding the quantification of codes (or themes) from our data [34]. In our findings, we report not only common themes (ie, those that emerged most frequently) but also salient themes (ie, themes reported by a minority that are still important).

We also specifically did not classify each patient based on whether they reported a high or low acceptance or use of the MERM. In contrast, we focus on reporting specific features of the MERM that were associated with higher or lower acceptance of the device because individual patients might find some components to be acceptable while simultaneously finding other components to be unacceptable. For example, a patient might appreciate the MERM's organization of medications but, at the same time, have concerns about the audible reminder because of fear that it could lead to the disclosure of her or his MDR-TB diagnosis. In addition, there is often individual variation in whether patients accept a particular technology [35]. As such, we avoid making a blanket declaration that the device is either *acceptable* or *unacceptable* to the larger patient population with MDR-TB in India.

Results

Characteristics of Study Participants

We interviewed 65 patients with MDR-TB, for whom the median age was 33 years (range 18-75 years).

Home visits were conducted for a median of 5 weeks (range 3-8 weeks) after the patients started using the MERM. Most patients were men, had some primary or secondary school education, and lived in the Chennai metropolitan area (Table 1).

We interviewed 10 HCPs, with a median age of 35 years (range 29-54 years). They had a median of 5.5 years of work experience in the NTEP (range 2-15 years). Most HCPs were men, had an undergraduate education, and had jobs as health visitors (Table 2).

Table 1. Descriptive statistics for patients with multidrug-resistant tuberculosis being monitored with the Medication Event Reminder Monitor.

Characteristic	Values, n (%)
Location	
Chennai	40 (62)
Mumbai	25 (38)
Gender	
Male	42 (65)
Female	23 (35)
Educational attainment	
No formal education or low literacy	13 (20)
Some primary or secondary education	44 (68)
Some college education, including degree or diploma holders	8 (12)
Occupation	
Unemployed	16 (25)
Student	7 (11)
Homemaker	7 (11)
Formal government or private sector job	6 (9)
Self-employed	29 (45)

Table 2. Descriptive statistics for health care providers who were interviewed to understand their acceptance of the Medication Event Reminder Monitor.

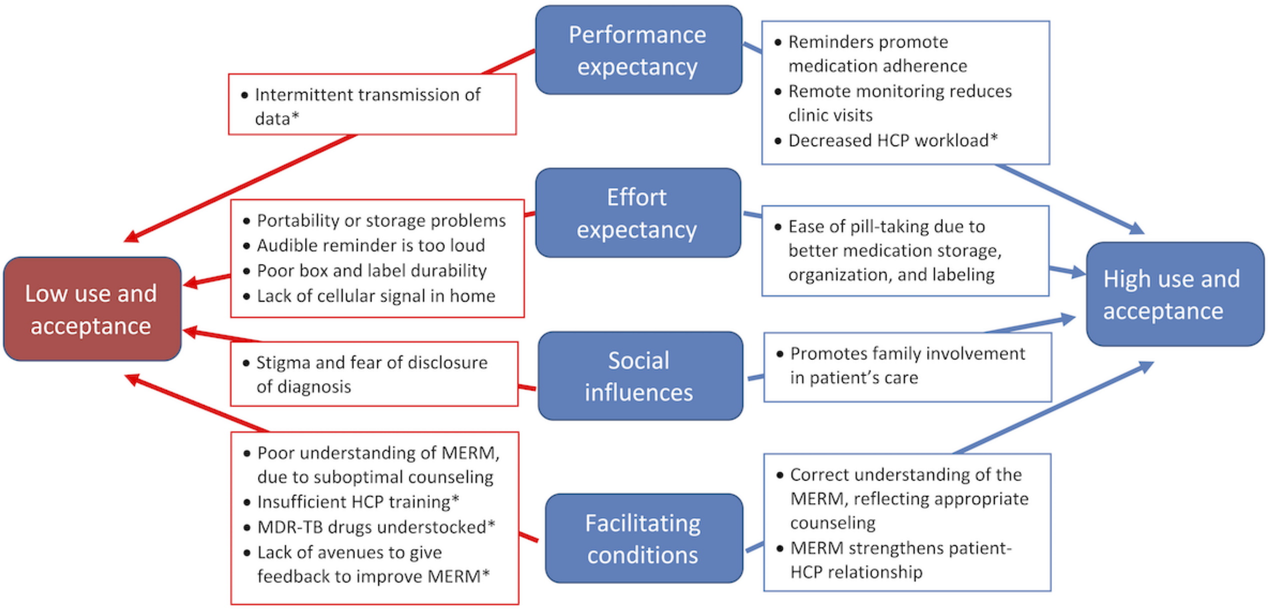
Characteristic	Values, n (%)
Location	
Chennai	5 (50)
Mumbai	5 (50)
Gender	
Male	6 (60)
Female	4 (40)
Educational attainment	
Undergraduate college education only	8 (80)
Postgraduate education	2 (20)
Designated position	
TB[a] health visitor	5 (50)
Senior treatment supervisor	2 (20)
Treatment coordinator	1 (10)
Deputy director of TB programs	1 (10)
District TB officer	1 (10)

[a] TB: tuberculosis.

Findings From Patients With MDR-TB

Interview findings revealed facilitators of and barriers to patient acceptance of MERM (Figure 2).

Figure 2. Key findings regarding determinants of high and low acceptance and use of the Medication Event Reminder Monitor by patients with MDR-TB and health care providers based on the Unified Theory of Acceptance and Use of Technology (UTAUT) framework. *Findings are from health care provider interviews; all other findings are from patient interviews. HCP: health care provider; MERM: Medication Event Reminder Monitor; MDR-TB: multidrug-resistant tuberculosis.



Facilitators of Patient Acceptance and Use of the MERM

Several factors were associated with a higher acceptance of MERM (Textbox 1).

With regard to performance expectancy (perceived usefulness), many patients felt that the reminders prevented them from forgetting to take their medications and helped them take it at the same time every day, with most preferring audible (Textbox 1; quote 1) and a few preferring visual (quote 2) reminders. One patient described the following benefits of audible reminders:

> *I finish my breakfast before 10 O' clock and wait for the alarm to ring. The alarm is useful because even when I forget it reminds me to take the tablets.* [Patient, male, 49 years]

Some patients also appreciated the yellow light, which served as a reminder to return to the clinic for medication refills.

Before being given their medications in the MERM, patients with MDR-TB usually visited clinics more frequently (eg, daily or weekly) for closer monitoring by HCPs. Some patients appreciated that remote monitoring of adherence resulted in reduced time and money spent on clinic visits (quote 3), as described by the following patient:

> *I previously had to visit the hospital three or four times in a month, but now I am going there once a month, so it is very good that you have provided this box. It is like a blessing for me.* [Patient, male, 39 years]

Several patients also appreciated the manner in which the MERM stores, organizes, and provides helpful internal labeling of medications. Patients previously stored medication blisters in plastic bags or a cardboard box provided by the TB program that did not have internal partitions to organize medications (quotes 4 and 5). These findings speak both to favorable perceived usefulness (ie, performance expectancy) of the MERM for organizing medications and favorable ease of use (ie, effort expectancy) because most patients found it easy to follow and understand the MERM's internal labels that guide pill taking, as noted by the following patient:

> *There are different compartments for each tablet, so they don't get mixed with each other....It is helpful. I like the arrows with the dots that explain how many of each medication I need to take.* [Patient, male, 38 years]

With regard to social influences, several patients reported that the audible reminder function promoted increased involvement of family members in their TB care (quotes 6-8), although occasionally such involvement was because of annoyance from the audible reminder (quote 7) or perceptions that the MERM facilitated government surveillance (quote 8). In some cases, however, this family involvement was prompted by the positive perception that the MERM represents an extension of the care provided by HCPs:

> *As soon as the alarm rings, my son immediately runs to me and says "Your doctor is calling you. Go and take your medicine and then do your work."* [Patient, male, 49 years]

Counseling of patients by HCPs in the appropriate use of the MERM is an important facilitating condition. The quality of counseling was assessed indirectly based on whether patients had appropriate or inappropriate knowledge of the functions of the MERM. There was variability in patient understanding of the MERM; however, most patients expressed a correct understanding of its medication labeling (quote 9) and other basic functions. For example, the following patient correctly interpreted the different lights on the MERM, reflecting appropriate counseling:

> *If red color [light] blinks there is no charge; if the green color blinks, it signals that the tablets have to be taken at 10 O'clock; and if the yellow color blinks, it means that the tablets are going to run out.* [Patient, male, 22 years]

Some patients reported that HCPs deployed the MERM in a manner that strengthened the patient-HCP relationship (quote 10). These findings highlight aspects of the MERM's perceived usefulness (performance expectancy), as well as favorable facilitating conditions in the health system's implementation of this technology. For example, patients appreciated that HCPs seemed to use the adherence data generated by the MERM in a positive manner, which resulted in patients feeling *caredfor* remotely:

> *If I don't open the [pill]box, somebody from the health center calls me to find out whether I have taken the tablets or not. They care for me.* [Patient, male, 61 years]

Textbox 1. Representative quotations on factors facilitating the acceptance of the Medication Event Reminder Monitor by patients based on constructs in the Unified Theory of Acceptance and Use of Technology (UTAUT).

Performance expectancy

- Reminders promote medication adherence
    - Quote 1: "The sound of the alarm forces me to take the medicine on time." (Patient, female, aged 42 years)
    - Quote 2: "Even when the alarm is not audible, the light is useful, especially when I am not near the box." (Patient, male, aged 25 years old)
- Remote monitoring reduces clinic visits
    - Quote 3: "I can do work at home properly now and do not have to worry about going to the health center." (Patient, female, aged 27 years)

Effort expectancy

- Ease of pill taking because of better medication storage, organization, and labeling
    - Quote 4: "Previously, I kept the tablets in a plastic cover, but now they are safer in the box. I used to be so confused, as there were so many medicines to take. Now it is easier." (Patient, female, aged 21 years)
    - Quote 5: "The pills were previously given in an ordinary cardboard tablet box, which does not have an alarm, but this box has an alarm to remind me." (Patient, male, aged 61 years old)

Social influences

- Promotes family involvement in patient's care
    - Quote 6: "When the alarm rings and I am outside my house, they send a person to inform me to take my pills." [Patient, male, aged 49 years]
    - Quote 7: "My mother complains when I delay taking the medicines. She would say, 'The box has been making noise constantly' and makes sure I take the medicines so the noise will stop." (Patient, male, aged 44 years)
    - Quote 8: "There is a camera in the box, so if you don't take the pills, people in Delhi will come to know. So take your pills." (Mother of a 25-year-old male patient)

Facilitating conditions

- Correct understanding of the Medication Event Reminder Monitor, reflecting appropriate counseling
    - Quote 9: "I take the tablets according to the dots shaded in each column above the compartment." (Patient, female, aged 28 years)
- Medication Event Reminder Monitor strengthens patient-health care provider relationship
    - Quote 10: "At the time of discharge...[a] health worker explained the MERM box and told me about the need to take my medicines regularly and that the box would help remind me. Those words motivated and encouraged me. My anxiety was reduced, and I was filled with happiness." (Patient, male, aged 53 years)

Barriers to Patient Acceptance and Use of the MERM

Patients also experienced barriers to the acceptance and use of MERM (Textbox 2).

A few patients admitted a lack of understanding regarding the purpose of the MERM, which suggests limitations in performance expectancy (perceived usefulness) for this minority of individuals:

> *I did not know that when I don't take pills, it will be shown [to HCPs] by a computer.* [Patient, female, aged 45 years]

More commonly, patients described limitations in effort expectancy (ease of use). For example, one patient described how the MERM's lack of portability led him to remove medication blisters from the device, which would result in the MERM recording adherence inaccurately:

> *I take my pills out of the box when I leave for work and put them in my pocket. I cannot carry such a big box to work that makes so much noise when I open it. I take the medicines [at work] when I am free. I do not benefit from the alarm or the light [audible and visual reminders] because I leave the box at home.* [Patient, male, aged 21 years]

Other patients similarly described how the device's large size served as a barrier to taking it to work (Textbox 2; quote 11) or storing it inside the house (quote 12). In addition, some of these concerns were related to stigma and privacy. Patients were concerned that the device's size and the loud sound of the audible reminder would draw attention and potentially raise questions about the patient's underlying medication condition (quote 13). Some patients found the audible reminder to be a major annoyance (quote 14). For example, one patient propped the box open to prevent the alarm from going off:

> *The alarm is too loud. So to avoid it [from going off], I put a paper in between the box and the lid and take the medicine.* [Patient, female, aged 18 years]

This action breaks the magnetic seal on the lid of the MERM and interferes with the recording of daily dose taking, thereby resulting in an inaccurate adherence record.

Patients also found that the body of the MERM, which is made of commercial-grade cardboard, poorly withstood the humidity and monsoon weather conditions in India. Medication labels peeled from the box (quote 15), and the box's shape became distorted:

> *My box bulged after it had rained continuously, and the inside of the house became damp, so it would be better if the box was made out of plastic.* [Patient male, aged 48 years]

A few patients found the MERM medication labels to be challenging; however, some of these difficulties may have reflected poor organization of medications in the box by HCPs, such that medications were not in the partitions corresponding to the appropriate labels:

> *This box is useful as there are instructions about its use, but sometimes the arrows [labeling each medication] don't match with the [correct] medicines. I get confused.* [Patient, male, aged 40 years]

Patients reported various other technical problems that limited the ease of using the MERM, including weak cellular signal in the patient's home resulting in the nonreporting of doses taken (quote 16), failure of the device's battery (quote 17), and malfunction of the reminder lights (quote 18).

With regard to social influences, several patients reported concerns regarding violations of the privacy and confidentiality of their MDR-TB diagnoses, which reflects the stigma surrounding this disease. As described earlier, patients were particularly concerned about stigma when traveling with the MERM (quote 13), but many patients were equally concerned about stigma when taking the device with them to clinic visits (quote 19), when friends or relatives visited the home (quote 20), or when family members heard the audible reminder, even if the device was hidden (quote 21). One patient was even concerned that the audible reminder was loud enough to draw the attention of her neighbors:

> *When the alarm rings, my neighbors can hear it. I am scared that they will come to know about my disease.* [Patient, female, aged 21 years]

With regard to facilitating conditions, some patients conveyed an incorrect understanding of the functions of the MERM (quote 22). For example, when asked about the visual reminders, the following patient conveyed an incorrect understanding:

> *I have to close the box when the yellow light blinks and I understand that if the red light blinks the tablets are going to be over.* [Patient, female, aged 75 years]

Textbox 2. Representative quotations on barriers to acceptance and use of the Medication Event Reminder Monitor by patients based on constructs in the Unified Theory of Acceptance and Use of Technology (UTAUT).

Effort expectancy

- Size, portability, and storage problems
  - Quote 11: “Sometimes I have to go for work for 2 or 3 days, and during that time I can’t carry this big box to the workplace. A smaller box with an alarm would be useful when I go for work.” (Patient, male, aged 41)
  - Quote 12: “I keep my box in a hen cage [outside of the house], because my children used to play with it. I don’t have a place in my home to keep the box where my children won’t reach it.” (Patient, male, aged 33 years)
  - Quote 13: “How can I carry this big box when I have to attend a marriage function in my village? I am sure my relatives will ask me questions when I take the medicines out of the box and when they hear the alarm sound. I usually take the medicines out of the box when I travel and leave the box at home.” (Patient, female, aged 45 years old)
- Audible reminder is too loud
  - Quote 14: “The sound is so loud, even the neighbors can hear it....Maybe it [the audible reminder] is useful for elders but not for youngsters like me because I feel irritated when it alerts me.” (Patient, female, aged 18 years)
- Limited durability of the box and labels
  - Quote 15: “The label in the box is not properly fixed and it has started peeling off.” (Patient, female, aged 18 years)
- Other technical problems with the Medication Event Reminder Monitor
  - Quote 16: “Sometimes due to [cellular] signal problems, although I was opening the box, these doses were not reported. I received calls from the health center [in which HCPs told me] to keep the box [at locations in the house] where the network might be better.” (Patient, male, aged 38 years)
  - Quote 17: “The alarm did not ring once and when I took it to the centre, they told that the box has run out of charge and needs to be replaced or recharged.” (Patient, male, aged 44 years)
  - Quote 18: “I am confused because all the three lights were glowing every day.” (Patient, female, aged 37 years)

Social influences

- Problems related to privacy and stigma
  - Quote 19: “When I carry the box when leaving the health center, people know that I have TB. This is embarrassing, so I try to hide it, but it is too big.” (Patient, male, aged 39 years)
  - Quote 20: “Suppose that my relatives visit my home. The box’s alarm could ring in front of everybody....They may come to know that I have this disease. I would be so embarrassed in front of them. So, I don’t like this box.” (Patient, female, aged 18 years)
  - Quote 21: “I keep the MERM inside the cupboard in my bedroom. I go inside my bedroom and take the medicine [privately]. If the alarm goes off and there is somebody at home, they sometimes ask me where that sound came from.” (Patient, male, aged 49 years)

Facilitating conditions

- Incorrect understanding of the Medication Event Reminder Monitor, reflecting suboptimal counseling
  - Quote 22: “The green color light helps me as a reminder but the red color means danger, which indicates that I have to go for the refill.” (Patient, female, aged 18 years old; description reflects incorrect understanding of the meaning of each light)

Findings From HCPs

The interview findings revealed both facilitators of and barriers to HCP acceptance of the MERM (Figure 2).

Facilitators of HCP Acceptance and Use of the MERM

For HCPs, perceptions of positive performance expectancy (ie, perceived usefulness) were the strongest facilitators of their acceptance and use of the MERM. In particular, most HCPs felt that remote monitoring of adherence was beneficial for both patients and HCPs. During the pilot implementation of the MERM, patients with MDR-TB were generally dispensed 1 month of medications. Clinic visits to pick up medications, which were previously required on a weekly or biweekly basis, were reduced substantially under the assumption that remote monitoring minimized the need for more frequent in-person monitoring. Most HCPs felt that patients benefited from this reduction in required clinic visits, as described in the following quotation:

> *Patients now do not have to travel long distances, spending their money to collect their drugs every week, or sometimes twice a week. Most of our MDR TB patients come from distant villages, and transportation is very difficult. We feel comfortable giving them a month’s supply in the box [MERM] as*

> *it is easier for them to take and the light and alarm [reminders] help them to take their drugs on time.* [Senior treatment supervisor]

Reduced frequency of patient visits decreased the workload for many HCPs, resulting in decreased stress:

> *Previously, we had to supervise therapy on a daily basis [ie, DOT]. But now the patients come [to the clinic] only once a month, so our work pressure has reduced.* [Senior treatment supervisor]

Some HCPs reported that the decreased workload allowed them to focus more on each patient interaction, as well as other tasks, which increased job satisfaction:

> *I have more time now to check whether patients have taken their tablets or not. I am also able to concentrate on other tasks as well, which gives me more satisfaction in my work.* [Health visitors]

With regard to social influences, some HCPs perceived that providing medications in the MERM, compared with the cardboard box previously used to dispense medications, was potentially less stigmatizing for patients because some cardboard boxes contained messages regarding TB:

> *The good thing about the MERM is that it does not carry any messages on TB [on the outside of the box], so there is no stigma attached to it. Patients can carry it freely.* [Medical officer]

Barriers to HCP Acceptance and Use of the MERM

HCPs also reported barriers to the acceptance of the MERM for both patients and HCPs. With regard to performance expectancy, HCPs found intermittent (every 72 h) updating of patients' adherence records to be the most significant limitation to the MERM's perceived usefulness, as described by the following HCP:

> *It takes 72 hours for the [MERM] dashboard to show that the patient has taken medications. This makes it difficult for us to monitor the patient's drug intake on a daily basis. We cannot take action as promptly and lose time.* [Pharmacist]

With regard to effort expectancy (ease of use), many HCPs felt that the size of the MERM made transporting the device to clinic visits prohibitive:

> *It is good [for patients] to have a device like the MERM, but...it is difficult to carry, as they need to go by bus and train. We need to provide them with a big bag for [the device].* [Health visitors]

The MERM's size also resulted in challenges for HCPs themselves:

> *Even for us [HCPs] at the health centers, it is difficult to find space to store these MERM boxes.* [Senior treatment supervisor]

Consistent with findings from the patient interviews, HCPs described a lack of cellular signal in patients' homes as a barrier to MERM use for rural patients in particular. This barrier contributed to limitations in HCPs' ability to obtain adherence data from and communicate with patients:

> *Some of the patients are not willing to use the box [MERM], as people living in the villages are not always getting [adequate cellular] signal, so the device is not working.* [Senior treatment supervisor]

During the pilot implementation of the MERM, HCPs found that some facilitating conditions on the part of the health system were suboptimal. For example, some HCPs felt that the single-day training provided would be insufficient for new personnel:

> *One day of training will be difficult if we have newly recruited staff, because they have to understand the [MDR TB] program, and then undergo training [in use of the MERM].* [Senior treatment supervisor]

Some barriers to implementation arose from more fundamental challenges in the MDR-TB program. For example, MDR-TB medications were supposed to be dispensed in the MERM on a monthly basis; however, some medications were sometimes understocked. This problem was easier to manage when patients were refilling medications on a weekly or more frequent basis because fewer medications had to be dispensed at any given visit. MERM implementation therefore worsened challenges related to the understocking of drugs:

> *Sometimes MDR-TB drugs are not available, and so we are not able to give all the medicines required....How do we leave that compartment [in the MERM for a specific medication] empty, and what can we tell the patient?* [Pharmacist]

Finally, some personnel felt that when problems were identified with the implementation of the MERM, they did not have channels to communicate these challenges:

> *When we started using the MERM, we were excited about the device. When patients came back for their medication refills, they raised concerns with regard to technical problems—the alarm, light, texture and size of the box, for example. I was not sure who to notify. Maybe we could have had those who made the device discuss our feedback so it could be improved?* [Senior treatment officer]

Discussion

Principal Findings

This study describes the evaluation of a low-cost digital pillbox aimed at promoting medication adherence among patients with MDR-TB during pilot implementation in India's NTEP. We find that acceptability of the MERM is variable; some features of the technology facilitate acceptability, whereas other features and contextual factors serve as barriers to engagement by some patients. Although previous studies have evaluated the use of similar digital pillboxes as part of TB care [3,18-20,22,23], to our knowledge, only one previous study conducted in South Africa [17] has evaluated the use of these technologies for patients with MDR-TB, who face unique challenges, including the complexity of their medication regimens, prolonged duration of therapy, increased risk of drug toxicities, and greater disease-related stigma. In addition, our study is unique in that it assessed the perspectives of both patients and HCPs.

Implications of Findings From Patients With MDR-TB

Multiple factors increase patient acceptance and the use of MERM. For most patients, audible and visual signals served as helpful reminders to take medications. Forgetfulness is a common barrier to adherence [36]. Although often thought of as a cognitive problem, forgetfulness may also reflect other life challenges faced by patients, such as depression or spending long hours at work. In a qualitative study in India assessing the acceptance of 99DOTS, a cellphone-based DAT used to monitor patients with drug-susceptible TB, most patients reported that SMS text messages were not useful reminders to take medications because these messages were lost amid a high volume of *spam* SMS text messages [35]. In contrast, an advantage of the MERM was that the reminders drew patients to the site where medications were stored. This increased the likelihood that patients with MDR-TB immediately took their doses, which may promote habit formation in pill taking behavior [37,38]. In addition, the MERM's reminders sometimes transformed household social dynamics by drawing family members into patients' TB care, a finding also reported in studies of other DATs [35].

Patients appreciated aspects of the MERM's design—in particular, the secure storage provided by the box, the labels to help patients take an appropriate number of tablets of each medication, and the organization of different medications facilitated by internal partitions. These features were valued in light of the complexity of MDR-TB treatment regimens, which include at least 4 or 5 different medications, as well as the fact that MDR-TB medications had previously been dispensed in a cardboard box without internal partitions to separate medications or labels to guide pill taking.

In some patients, MERM enhanced their relationship with the health system. Most patients appreciated saving time and money by not having to visit the clinic as often because the frequency of routine clinic visits for patients with MDR-TB was reduced during the MERM pilot implementation. Although this resulted in decreased face-to-face interactions with HCPs, some patients actually described feeling more *cared for*. This feeling was derived from the perception that HCPs were remotely watching over their clinical progress, as well as from positive responses to actual phone or in-person outreach by HCPs, guided by patients' adherence data. Previous studies evaluating the use of DATs to support HIV and TB treatment adherence in Uganda, India, and South Africa similarly found that remote monitoring enhanced some patients' perceptions of the care provided by the health system [17,35]. This may be one of the behavioral pathways by which DATs may motivate patients to adhere to treatment.

Patients also reported barriers to the acceptance and use of MERM. Some of these barriers may be modified by altering the MERM's design or implementation (Textbox 3). For example, the loud volume of the audible reminder—a common complaint from patients also reported in a previous study of the MERM from China [18]—could potentially be modified or the audible reminder disabled completely, ideally based on patients' personal preferences. As another example, during pilot implementation, the MERM was made of commercial-grade cardboard, which did not wear well in India's humid weather conditions. Redesigning the MERM using plastic would be feasible and minimize weather-related damage, although it would likely increase the cost of the device. Other technical problems, such as battery failure or inappropriate blinking of the reminder lights, could likely be addressed with product improvements in future iterations of the MERM.

Textbox 3. Recommendations for improving the Medication Event Reminder Monitor device and its implementation, based on findings from patients with multidrug-resistant tuberculosis and health care provider interviews.

Design of the device

- Data transmission from the device on a daily basis may facilitate better near real-time monitoring
- Redesign using plastic (rather than cardboard) may reduce wear because of weather conditions
- Strengthening internal partitions may help avoid accidental mixing of different medications
- Device reuse should be limited, given considerable wear and tear even after single patient use

Reminder functions

- Allowing patients to reduce the volume of the audible reminder or to deactivate audible or visual reminders may address concerns about privacy and stigma
- Malfunction of visual reminders (eg, all lights blinking at once) should be fixed

Counseling and monitoring of patients

- Health care providers should be trained to provide standardized counseling to ensure patient understanding of key Medication Event Reminder Monitor functions
- Health care providers should use pill counts and ask adherence questions to patients at clinic and home visits to cross-check the Medication Event Reminder Monitor's adherence data

Screening out patients for whom the Medication Event Reminder Monitor may not be appropriate

- Systematic screening should be performed upfront to identify patients for whom the Medication Event Reminder Monitor may not be appropriate, including those with concerns about stigma, a fear of disclosure of diagnosis, difficulties with portability, and lack of cellular signal in the home

Training of health care providers

- Mechanisms should be created for the training of newly hired National Tuberculosis Elimination Program personnel and provision of periodic refresher training in the Medication Event Reminder Monitor for existing personnel
- Mechanisms should be created for National Tuberculosis Elimination Program personnel to provide ongoing feedback to facilitate device improvements

Although some barriers may be addressable, others may present more fundamental challenges that could limit the use of the MERM among some patients. For example, the MERM's large size was a barrier for patients who were traveling or who preferred to take their medications at work. However, the MERM's size is necessary to hold a 1-month supply of MDR-TB medications, and patients benefit from having their medications dispensed in an organized manner with appropriate labeling. Furthermore, because of its prohibitive size, patients who need to take doses when traveling or at work tend to remove doses from the device rather than carry it with them. The lack of cellular signals in the home is another nonmodifiable barrier that limits the benefits of remote monitoring because adherence data cannot be transmitted from the device on a regular basis.

Disease-related stigma, from family and community members, is a common challenge faced by patients with MDR-TB [10]. Owing to stigma, patients often do not disclose their diagnosis to family members, friends, and coworkers; as a result, patients fear situations that could result in disclosure of their diagnosis to others. The MERM's large size, as well as its audible and visual reminders, raised patient concerns regarding the risk of disclosure of diagnosis. Indeed, as ascertained from their interviews, patients who faced barriers related to their social context, including stigma, a fear of disclosure, or frequent work-related travel, seemed to be the most likely to not use the MERM.

All of these problems—removal of doses from the device because of its lack of portability, the nonreporting of box openings because of lack of cellular signal, and the nonuse of the device because of disease-related stigma—could result in underreporting of medication doses, resulting in inaccuracies in patients' adherence records. Recent studies found that these same barriers contribute to relatively high rates of patient nonengagement with 99DOTS in India [35], especially in the continuation phase of therapy, which contributed to the technology's suboptimal accuracy for measuring adherence to TB medications [24]. A qualitative study of drug-susceptible TB patients monitored with the MERM in Vietnam found that only approximately half of the patients used the device as intended largely because of the difficulties with the device's portability, with the result that the MERM data often did not reflect actual adherence [23]. A high rate of device nonuse was also found in a study that used the Wisepill device (a similar digital pillbox) to monitor adherence to HIV preexposure prophylaxis in young men who have sex with men in the United States [39].

These barriers to use suggest that if implementation of the MERM is expanded among patients with MDR-TB in India, there could be limits to the device's reach, or overall coverage, in this patient population. Wide-scale implementation of a similar digital pillbox among patients with drug-susceptible TB in China has revealed meaningful limitations in the reach of the device [21,22]. For example, in one study of the implementation of a digital pillbox in 30 counties in China, even after excluding 41% of the patient cohort who were either not eligible to use the device or who did not receive the device for unclear reasons, only approximately two-thirds of the remaining 1314 patients who received the pillbox had sustained use for the remainder of the treatment [22]. The other one-third of patients who received the digital pillbox either stopped using the device or met the criteria to be shifted back to monitoring with DOT because of a high proportion of missed doses. These missed doses could have represented either true nonadherence to medications or inappropriate use of the device.

In light of such findings from other contexts, it would be reasonable to assume that some proportion of patients with MDR-TB in India might not use the MERM in wide-scale implementation. On the basis of our findings, the NTEP could consider screening patients with MDR-TB upfront to identify individuals who might be unlikely to use the device, for example, because of patient concerns about stigma and portability or lack of cellular signal in the home (Textbox 3). In addition, HCPs should use other strategies to verify medication adherence, including pill counts and asking adherence-related questions to patients at every in-person clinic and home visit, which will help HCPs to cross-check the adherence data being received from the MERM.

Implications of Findings From HCPs

In the HCP interviews, NTEP personnel affirmed some of the patient-oriented benefits of the MERM, in particular, the time and money saved by patients from the reduced frequency of clinic visits; however, HCPs' perceptions that the MERM was associated with fewer patient concerns about stigma were not shared by some patients. HCPs also reported that implementation of the MERM reduced their workload because of the reduced frequency of clinic visits by patients and the ability to monitor adherence from the clinic rather than by home visits. As a result, HCPs dedicated greater time to other tasks and reported improved job satisfaction, similar to the findings of a previous study of the MERM conducted in China [18]. HCPs did find some aspects of the pilot implementation to be suboptimal; however, most of these concerns were potentially addressable. In particular, they reported a need for more training in the use of the MERM, especially in light of the high turnover of staff, and the need for a platform to communicate any implementation challenges they faced (Textbox 3).

Directions for Future Research

This initial evaluation has identified several features that may facilitate high acceptability of the MERM for many patients with MDR-TB, especially if modifications are made to improve the device. Future research should focus on understanding how often patients face critical barriers to acceptability (eg, disease-related stigma), the extent to which these barriers lead to device nonuse and whether screening for these barriers can be used to better target the MERM to patients most likely to use it. In addition, further research is needed to understand the accuracy of the MERM for measuring adherence to MDR-TB medications, its effectiveness for improving treatment outcomes, and its reach, that is, coverage or uptake by patients, in large-scale implementation [3].

Even for patients who agree to use the MERM, the adherence record could be inaccurate because of underreporting (eg, if patients take medications out of the device, resulting in device nonuse) or overreporting (eg, if patients open and close the device without actually taking medications). Indeed, a recent study of 99DOTS, in which its adherence record was compared with urine isoniazid test results from patients with TB collected during unannounced home visits, found that both under- and overreporting of adherence contributed to the technology's suboptimal accuracy [24]. A similar research approach, involving unannounced home visits with measurement of urine biomarkers for MDR-TB medications, could be used to evaluate the accuracy of MERM, although pill counts should also be conducted to provide insights into whether patients have differential adherence to different medications in the MDR-TB regimen.

Existing studies of the use of DATs to promote adherence to TB medications have found both positive [20,40] and negative or equivocal [41-43] impacts on adherence and TB treatment outcomes. As such, studies of effectiveness, especially high-quality randomized trials, are needed to assess whether MERM use translates into improvements in treatment outcomes and recurrence-free survival for patients with MDR-TB. Even when DATs have been shown to be effective, as with digital pillboxes in China [20], subsequent large-scale implementation studies have shown suboptimal reach or coverage of patients [21,22]. As such, studies of the MERM's coverage of patients with MDR-TB in large-scale implementation will be critical to ensure that it achieves population-level impact. Finally, in light of the diverse psychosocial barriers to adherence faced by patients with MDR-TB [10], the benefits of monitoring with the MERM in this population will depend on the development of interventions to address problems, such as medication toxicities, depression, stigma, and substance use disorder, which are often the underlying causes of nonadherence [44].

Study Limitations

Our study was limited to assessing patient and HCP perceptions, rather than more objective findings, such as the accuracy of the MERM or impact on clinical outcomes. As such, we may have overestimated the acceptability and benefits of this technology because of socially desirable responses, which is a common bias in qualitative research. In addition, patients attributed the reduced frequency of their clinic visits to the MERM, as a longer supply of medications was dispensed in the device. The reduced frequency of clinic visits may have therefore biased patients in favor of higher acceptance of the device; however, provision of a longer supply of medications could have just as easily been implemented without the MERM.

Our deductive approach to analysis allowed us to organize and report our findings using the UTAUT, which is a robust and evidence-based framework for understanding technology acceptance; however, a limitation of this approach is that we could have overlooked findings that did not fit into this predetermined framework.

Another limitation of our study is that we assessed patients' perceptions of the MERM within a few weeks of their use of the device. In light of the prolonged duration of MDR-TB treatment, it is possible that patients' acceptance and use of the device could change over time. In addition, our study was limited to 2 cities and may not be representative of barriers to the use of the MERM in rural parts of India. Future studies could consider including diverse geographic settings and conducting multiple interviews to understand the acceptability of the MERM throughout the treatment course.

Conclusions

In this study of the pilot implementation of a low-cost digital pillbox to promote adherence to MDR-TB medications, we identified several features that facilitate high acceptability of the device among patients. These included helpful organization and labeling of medications, feeling more *cared for* by the health system because of remote monitoring, and appreciation of the audible and visual reminders, which often drew family members into patients' care.

At the same time, we identified barriers that could limit the acceptance and use of the MERM by some patients. Although some of these barriers could be addressed relatively easily with modification of the device, other barriers—such as difficulties with the device's portability, lack of cellular signal in the home, and a fear of disclosure of diagnosis because of disease-related stigma—are more difficult to modify and may limit the reach or population coverage of this technology. Future research is needed to assess the accuracy of the MERM for measuring adherence, its effectiveness for improving treatment outcomes, and patients' sustained use of the device in larger scale implementation in India's MDR-TB treatment program.

Acknowledgments

Maya Lubeck-Schricker assisted with referencing and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Examples of questions included in the in-depth interview guide for patients with multidrug-resistant tuberculosis in relation to constructs in the Unified Theory of Acceptance and Use of Technology (UTAUT).

[DOC File , 39 KB - jmir_v23i6e23294_app1.doc]

References

1. Global Tuberculosis Report. World Health Organization. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/329368/9789241565714-eng.pdf?ua=1 [accessed 2020-08-04]
2. Collaborative Group for the Meta-Analysis of Individual Patient Data in MDR-TB treatment–2017, Ahmad N, Ahuja SD, Akkerman OW, Alffenaar JC, Anderson LF, et al. Treatment correlates of successful outcomes in pulmonary multidrug-resistant tuberculosis: an individual patient data meta-analysis. Lancet 2018 Sep 8;392(10150):821-834 [FREE Full text] [doi: 10.1016/S0140-6736(18)31644-1] [Medline: 30215381]
3. Subbaraman R, de Mondesert L, Musiimenta A, Pai M, Mayer KH, Thomas BE, et al. Digital adherence technologies for the management of tuberculosis therapy: mapping the landscape and research priorities. BMJ Glob Health 2018;3(5):e001018 [FREE Full text] [doi: 10.1136/bmjgh-2018-001018] [Medline: 30364330]
4. Lan Z, Ahmad N, Baghaei P, Barkane L, Benedetti A, Brode SK, Collaborative Group for the Meta-Analysis of Individual Patient Data in MDR-TB treatment 2017. Drug-associated adverse events in the treatment of multidrug-resistant tuberculosis: an individual patient data meta-analysis. Lancet Respir Med 2020 Apr;8(4):383-394 [FREE Full text] [doi: 10.1016/S2213-2600(20)30047-3] [Medline: 32192585]
5. Thomas B, Watson B, Senthil EK, Deepalakshmi A, Balaji G, Chandra S, et al. Alcohol intervention strategy among tuberculosis patients: a pilot study from South India. Int J Tuberc Lung Dis 2017 Aug 1;21(8):947-952. [doi: 10.5588/ijtld.16.0693] [Medline: 28786805]
6. van Rensburg A, Dube A, Curran R, Ambaw F, Murdoch J, Bachmann M, et al. Comorbidities between tuberculosis and common mental disorders: a scoping review of epidemiological patterns and person-centred care interventions from low-to-middle income and BRICS countries. Infect Dis Poverty 2020 Jan 15;9(1):4 [FREE Full text] [doi: 10.1186/s40249-019-0619-4] [Medline: 31941551]
7. Craig GM, Daftary A, Engel N, O'Driscoll S, Ioannaki A. Tuberculosis stigma as a social determinant of health: a systematic mapping review of research in low incidence countries. Int J Infect Dis 2017 Mar;56:90-100 [FREE Full text] [doi: 10.1016/j.ijid.2016.10.011] [Medline: 27810521]
8. Cazabon D, Pande T, Sen P, Daftary A, Arsenault C, Bhatnagar H, et al. User experience and patient satisfaction with tuberculosis care in low- and middle-income countries: A systematic review. J Clin Tuberc Other Mycobact Dis 2020 May;19:100154 [FREE Full text] [doi: 10.1016/j.jctube.2020.100154] [Medline: 32140571]

9. Bhatnagar H. User-experience and patient satisfaction with quality of tuberculosis care in India: a mixed-methods literature review. J Clin Tuberc Other Mycobact Dis 2019 Dec;17:100127 [FREE Full text] [doi: 10.1016/j.jctube.2019.100127] [Medline: 31788569]
10. Thomas BE, Shanmugam P, Malaisamy M, Ovung S, Suresh C, Subbaraman R, et al. Psycho-socio-economic issues challenging multidrug resistant tuberculosis patients: a systematic review. PLoS One 2016;11(1):e0147397 [FREE Full text] [doi: 10.1371/journal.pone.0147397] [Medline: 26807933]
11. Karumbi J, Garner P. Directly observed therapy for treating tuberculosis. Cochrane Database Syst Rev 2015 May 29(5):CD003343 [FREE Full text] [doi: 10.1002/14651858.CD003343.pub4] [Medline: 26022367]
12. Tian J, Lu Z, Bachmann MO, Song F. Effectiveness of directly observed treatment of tuberculosis: a systematic review of controlled studies. Int J Tuberc Lung Dis 2014 Sep;18(9):1092-1098. [doi: 10.5588/ijtld.13.0867] [Medline: 25189558]
13. Pasipanodya JG, Gumbo T. A meta-analysis of self-administered vs directly observed therapy effect on microbiologic failure, relapse, and acquired drug resistance in tuberculosis patients. Clin Infect Dis 2013 Jul;57(1):21-31 [FREE Full text] [doi: 10.1093/cid/cit167] [Medline: 23487389]
14. Sagbakken M, Frich JC, Bjune GA, Porter JD. Ethical aspects of directly observed treatment for tuberculosis: a cross-cultural comparison. BMC Med Ethics 2013 Jul 2;14:25 [FREE Full text] [doi: 10.1186/1472-6939-14-25] [Medline: 23819555]
15. Yellappa V, Lefèvre P, Battaglioli T, Narayanan D, van der Stuyft P. Coping with tuberculosis and directly observed treatment: a qualitative study among patients from South India. BMC Health Serv Res 2016 Jul 19;16:283 [FREE Full text] [doi: 10.1186/s12913-016-1545-9] [Medline: 27430557]
16. WHO Consolidated Guidelines on Drug-resistant Tuberculosis Treatment. World Health Organization. URL: http://www.who.int/tb/publications/2019/consolidated-guidelines-drug-resistant-TB-treatment/en/ [accessed 2020-08-04]
17. Bionghi N, Daftary A, Maharaj B, Msibi Z, Amico KR, Friedland G, et al. Pilot evaluation of a second-generation electronic pill box for adherence to Bedaquiline and antiretroviral therapy in drug-resistant TB/HIV co-infected patients in KwaZulu-Natal, South Africa. BMC Infect Dis 2018 Apr 11;18(1):171 [FREE Full text] [doi: 10.1186/s12879-018-3080-2] [Medline: 29642874]
18. Liu X, Blaschke T, Thomas B, De Geest S, Jiang S, Gao Y, et al. Usability of a medication event reminder monitor system (MERM) by providers and patients to improve adherence in the management of tuberculosis. Int J Environ Res Public Health 2017 Sep 25;14(10):1115 [FREE Full text] [doi: 10.3390/ijerph14101115] [Medline: 28946683]
19. Musiimenta A, Tumuhimbise W, Mugaba AT, Muzoora C, Armstrong-Hough M, Bangsberg D, et al. Digital monitoring technologies could enhance tuberculosis medication adherence in Uganda: Mixed methods study. J Clin Tuberc Other Mycobact Dis 2019 Dec;17:100119 [FREE Full text] [doi: 10.1016/j.jctube.2019.100119] [Medline: 31788561]
20. Liu X, Lewis JJ, Zhang H, Lu W, Zhang S, Zheng G, et al. Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial. PLoS Med 2015 Sep;12(9):e1001876 [FREE Full text] [doi: 10.1371/journal.pmed.1001876] [Medline: 26372470]
21. Wang N, Zhang H, Zhou Y, Jiang H, Dai B, Sun M, et al. Using electronic medication monitoring to guide differential management of tuberculosis patients at the community level in China. BMC Infect Dis 2019 Oct 15;19(1):844 [FREE Full text] [doi: 10.1186/s12879-019-4521-2] [Medline: 31615433]
22. Wang N, Shewade HD, Thekkur P, Huang F, Yuan Y, Wang X, et al. Electronic medication monitor for people with tuberculosis: implementation experience from thirty counties in China. PLoS One 2020;15(4):e0232337 [FREE Full text] [doi: 10.1371/journal.pone.0232337] [Medline: 32348351]
23. Drabarek D, Anh NT, Nhung NV, Hoa NB, Fox GJ, Bernays S. Implementation of medication event reminder monitors among patients diagnosed with drug susceptible tuberculosis in rural Vietnam: a qualitative study. PLoS One 2019;14(7):e0219891 [FREE Full text] [doi: 10.1371/journal.pone.0219891] [Medline: 31329610]
24. Thomas BE, Kumar JV, Chiranjeevi M, Shah D, Khandewale A, Thiruvengadam K, et al. Evaluation of the accuracy of 99DOTS, a novel cellphone-based strategy for monitoring adherence to tuberculosis medications: comparison of digital adherence data with urine isoniazid testing. Clin Infect Dis 2020 Dec 3;71(9):e513-e516 [FREE Full text] [doi: 10.1093/cid/ciaa333] [Medline: 32221550]
25. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
26. Venkatesh V, Thong J, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016 May;17(5):328-376 [FREE Full text] [doi: 10.18411/d-2016-154]
27. Dhanaraj B, Papanna MK, Adinarayanan S, Vedachalam C, Sundaram V, Shanmugam S, et al. Prevalence and risk factors for adult pulmonary tuberculosis in a metropolitan city of South India. PLoS One 2015;10(4):e0124260 [FREE Full text] [doi: 10.1371/journal.pone.0124260] [Medline: 25905900]
28. India TB Report 2019: Revised National TB Control Programme Annual Report. Central TB Division, Ministry of Health and Family Welfare. 2019. URL: https://tbcindia.gov.in/WriteReadData/India%20TB%20Report%202019.pdf [accessed 2020-08-04]
29. D'souza DT, Mistry NF, Vira TS, Dholakia Y, Hoffner S, Pasvol G, et al. High levels of multidrug resistant tuberculosis in new and treatment-failure patients from the Revised National Tuberculosis Control Programme in an urban metropolis

(Mumbai) in Western India. BMC Public Health 2009 Jun 29;9:211 [FREE Full text] [doi: 10.1186/1471-2458-9-211] [Medline: 19563647]

30. Isaakidis P, Das M, Kumar AM, Peskett C, Khetarpal M, Bamne A, et al. Alarming levels of drug-resistant tuberculosis in HIV-infected patients in metropolitan Mumbai, India. PLoS One 2014;9(10):e110461 [FREE Full text] [doi: 10.1371/journal.pone.0110461] [Medline: 25333696]
31. Dalal A, Pawaskar A, Das M, Desai R, Prabhudesai P, Chhajed P, et al. Resistance patterns among multidrug-resistant tuberculosis patients in greater metropolitan Mumbai: trends over time. PLoS One 2015;10(1):e0116798 [FREE Full text] [doi: 10.1371/journal.pone.0116798] [Medline: 25606853]
32. Satyanarayana S, Nair SA, Chadha SS, Shivashankar R, Sharma G, Yadav S, et al. From where are tuberculosis patients accessing treatment in India? Results from a cross-sectional community based survey of 30 districts. PLoS One 2011;6(9):e24160 [FREE Full text] [doi: 10.1371/journal.pone.0024160] [Medline: 21912669]
33. Revised National TB Control Programme: Technical and Operational Guidelines for Tuberculosis Control. Central TB Division, Ministry of Health and Family Welfare. 2016. URL: http://health.bih.nic.in/Docs/Guidelines/Guidelines-TB-Control.pdf [accessed 2020-08-04]
34. Patton M. Qualitative Research & Evaluation Methods. Thousand Oaks, CA: Sage Publications; 2002.
35. Thomas B, Kumar J, Onongaya C, Bhatt S, Galivanche A, Periyasamy M, et al. Explaining differences in the acceptability of 99dots, a cell phone-based strategy for monitoring adherence to tuberculosis medications: qualitative study of patients and health care providers. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e16634 [FREE Full text] [doi: 10.2196/16634] [Medline: 32735220]
36. Shubber Z, Mills EJ, Nachega JB, Vreeman R, Freitas M, Bock P, et al. Patient-reported barriers to adherence to antiretroviral therapy: a systematic review and meta-analysis. PLoS Med 2016 Nov;13(11):e1002183 [FREE Full text] [doi: 10.1371/journal.pmed.1002183] [Medline: 27898679]
37. Ware NC, Pisarski EE, Tam M, Wyatt MA, Atukunda E, Musiimenta A, et al. The meanings in the messages: how SMS reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016 May 15;30(8):1287-1294 [FREE Full text] [doi: 10.1097/QAD.0000000000001035] [Medline: 26807967]
38. Haberer JE, Musiimenta A, Atukunda EC, Musinguzi N, Wyatt MA, Ware NC, et al. Short message service (SMS) reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016 May 15;30(8):1295-1300 [FREE Full text] [doi: 10.1097/QAD.0000000000001021] [Medline: 26760452]
39. Koss CA, Hosek SG, Bacchetti P, Anderson PL, Liu AY, Horng H, et al. Comparison of measures of adherence to human immunodeficiency virus preexposure prophylaxis among adolescent and young men who have sex with men in the United States. Clin Infect Dis 2018 Jan 6;66(2):213-219 [FREE Full text] [doi: 10.1093/cid/cix755] [Medline: 29020194]
40. Yoeli E, Rathauser J, Bhanot SP, Kimenye MK, Mailu E, Masini E, et al. Digital health support in treatment for tuberculosis. N Engl J Med 2019 Sep 5;381(10):986-987. [doi: 10.1056/NEJMc1806550] [Medline: 31483974]
41. Thekkur P, Kumar AN, Chinnakali P, Selvaraju S, Bairy R, Singh AR, et al. Outcomes and implementation challenges of using daily treatment regimens with an innovative adherence support tool among HIV-infected tuberculosis patients in Karnataka, India: a mixed-methods study. Glob Health Action 2019;12(1):1568826 [FREE Full text] [doi: 10.1080/16549716.2019.1568826] [Medline: 30712507]
42. Iribarren S, Beck S, Pearce PF, Chirico C, Etchevarria M, Cardinale D, et al. TextTB: a mixed method pilot study evaluating acceptance, feasibility, and exploring initial efficacy of a text messaging intervention to support tb treatment adherence. Tuberc Res Treat 2013;2013:349394 [FREE Full text] [doi: 10.1155/2013/349394] [Medline: 24455238]
43. Mohammed S, Glennerster R, Khan AJ. Impact of a daily SMS medication reminder system on tuberculosis treatment outcomes: a randomized controlled trial. PLoS One 2016;11(11):e0162944 [FREE Full text] [doi: 10.1371/journal.pone.0162944] [Medline: 27802283]
44. Haberer JE, Subbaraman R. Digital technology for tuberculosis medication adherence: promise and peril. Ann Am Thorac Soc 2020 Apr;17(4):421-423. [doi: 10.1513/AnnalsATS.202001-027ED] [Medline: 32233860]

Abbreviations

DAT: digital adherence technology
DOT: directly observed therapy
HCP: health care provider
MDR-TB: multidrug-resistant tuberculosis
MERM: Medication Event Reminder Monitor
NTEP: National Tuberculosis Elimination Program
TB: tuberculosis
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by R Kukafka; submitted 06.08.20; peer-reviewed by L Rabinovich, A Kibballi Madhukeshwar, P Akande; comments to author 20.10.20; revised version received 27.10.20; accepted 30.04.21; published 10.06.21.

Please cite as:
Thomas BE, Kumar JV, Periyasamy M, Khandewale AS, Hephzibah Mercy J, Raj EM, Kokila S, Walgude AS, Gaurkhede GR, Kumbhar JD, Ovung S, Paul M, Rajkumar BS, Subbaraman R
Acceptability of the Medication Event Reminder Monitor for Promoting Adherence to Multidrug-Resistant Tuberculosis Therapy in Two Indian Cities: Qualitative Study of Patients and Health Care Providers
J Med Internet Res 2021;23(6):e23294
URL: https://www.jmir.org/2021/6/e23294
doi:10.2196/23294
PMID:34110300

Original Paper

Effect of Textual Features on the Success of Medical Crowdfunding: Model Development and Econometric Analysis from the Tencent Charity Platform

Fuguo Zhang[1], Prof Dr; Bingyu Xue[1], MA; Yiran Li[2], PhD; Hui Li[1], MA; Qihua Liu[3], Prof Dr

[1]School of Information Management, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China
[2]School of Management, Zhejiang University of Technology, Hangzhou, Zhejiang, China
[3]School of Management, Hainan University, Haikou, Hainan, China

Corresponding Author:
Qihua Liu, Prof Dr
School of Management
Hainan University
58 Renmin Avenue
Haikou, Hainan, 570228
China
Phone: 86 0898 66290798
Email: qh_liu@163.com

Abstract

Background: Medical crowdfunding utilizes the internet to raise medical funds. Medical crowdfunding has developed rapidly worldwide; however, most medical crowdfunding projects fail to raise the targeted funds. Therefore, a very important research problem that has not received sufficient attention from the existing literature is identifying which factors affect the success of medical crowdfunding projects.

Objective: The aim of this study was to examine the effect of textual features of medical crowdfunding projects on their success rate using 4903 real projects from the Tencent Charity platform, a well-known medical crowdfunding platform in China. In particular, according to Aristotle's theory of persuasion, we divided the project text of medical crowdfunding into the project title and project details, which were analyzed from two perspectives (existence and extent) to explore their respective impacts.

Methods: We established a research framework to meet our research goals. The process was divided into five main parts. We first collected data from Tencent Charity using Python programs and cleaned the datasets. Second, we selected variables and built the research model based on previous studies and the theory of persuasion. Next, the selected variables were extracted from the project text. We then performed econometric analysis using multiple regression analysis. Finally, we evaluated the results of econometric analysis to extract knowledge.

Results: In the project title, the presence of the patient's disease (*P*=.04) and occupation (*P*=.01) had a positive impact on the success rate of fundraising, whereas the presence of age (*P*<.001), gender (*P*=.001), and negative emotions (*P*=.04) had a negative impact. In the project details, the presence of the patient's occupation (*P*=.01), monetary evidence (*P*=.02), and negative emotions (*P*=.04) played a positive role in the fundraising success rate, whereas the presence of age (*P*<.001) and positive emotions (*P*<.001) played a negative role. Moreover, in the project details, high-frequency monetary evidence (*P*=.02) and negative words (*P*=.02), as well as a short narrative length (*P*=.01) were conducive to succeeding in medical crowdfunding. Younger patients were more likely to obtain a higher success rate in medical crowdfunding. For patients whose occupations were national civil servant, professional skill worker, clerk, business and service worker, solider, child, student, and public-spirited person, the success rate of fundraising decreased sequentially.

Conclusions: This study collected 4903 valid data from Tencent Charity, and identified which factors in the project text play an important role in the success rate of medical crowdfunding from the perspective of existence and extent. We found that in addition to the project details, the features of the project title also have an important impact on the success rate of fundraising. These findings provide important theoretical and managerial implications for medical crowdfunding.

(J Med Internet Res 2021;23(6):e22395) doi:10.2196/22395

KEYWORDS

medical crowdfunding; textual features; project title; project details; fundraising success; theory of persuasion

Introduction

Background

With an aging population, the widespread prevalence of chronic diseases, a growing number of serious illnesses, and advances in medical technology, global health spending is rising rapidly. From 2018 to 2022, global health spending is expected to increase at an annual rate of 5.4%, representing a significant increase from 2.9% in 2013-2017 [1]. However, access to good health care such as through medical insurance and government subsidies is not available to all. Currently, half of the world's population still lacks access to essential health services [2]. Additionally, even with partial health care, many medical conditions still incur high out-of-pocket medical costs, leading to significant financial stress on individuals and their families. In 2018, approximately 100 million people worldwide fell into extreme poverty because of out-of-pocket medical expenses [2].

In recent years, medical expenses in China have also increased rapidly. According to statistics from the National Health Commission, China's medical expenses exceeded US $97.4 billion in 2019 [3]. Although China has established a relatively complete medical security system, many patients still have to pay large medical expenses themselves. China has established a multilevel medical security system with a basic medical insurance system as the main body, commercial insurance as a supplement, and social charity relief as a base. China has a basic medical insurance system, which receives subsidies for poor patients so they can be covered by insurance. As of January 2020, China's basic medical insurance system covers more than 1.35 billion people, a coverage rate of approximately 97%, according to the National Healthcare Security Administration [4]. There are few people who currently purchase commercial insurance. As of October 2019, the number of people with major illness insurance policies in China did not exceed 100 million [5]. However, the current medical insurance system still has the following three problems. First, medical insurance reimbursement is restrictive, with reimbursements only provided for the hospital's medical expenses, excluding nutrition and other expenses. Moreover, the expenses above this standard can be reimbursed only when the medical expenses reach the reimbursement standard. The reimbursement ratio ranges from 30% to 90%, and the maximum reimbursement limit is approximately US $40,000 [6]. Second, some drugs are not covered by reimbursement. For example, Neulasta/Peglasta is an imported drug that is effective in the treatment of cancer, but it is not on the reimbursement list. Finally, there are still more than 40 million Chinese people without medical insurance. Consequently, the existence of the above-mentioned problems has caused out-of-pocket medical expenses to remain high for many patients. There are still many families who suffer from poverty due to illness. As an example, Sohu News revealed that an 8-year-old girl in Tianjin, China was diagnosed with malignant neuroblastoma. Because of this illness, the family has spent more than US $60,000 and has accrued approximately US $30,000 of debt as a result. To cure this disease, follow-up treatment costs will reach nearly US $60,000 [7].

A unique and increasingly popular type of crowdfunding application, namely medical crowdfunding, is being used by many patients and their families to pay for medical bills. Medical crowdfunding has developed rapidly worldwide. GoFundMe, a US crowdfunding platform, raises US $650 million from approximately 250,000 medical crowdfunding campaigns each year [8]. A medical crowdfunding platform in China, Shuidichou, successfully provided fundraising services for more than 800,000 patients in economic distress as of the end of September 2018, raised more than US $1.4 billion, and aided over 340 million patients [9]. Other medical crowdfunding platforms have also grown rapidly in China. It is expected that medical crowdfunding worldwide will experience an annual growth rate of 25% [10]. However, the success rate is low for medical crowdfunding projects, as most of these campaigns fail to raise the target funds [11,12]. Berliner and Kenworthy [13] found that only 10% of the 200 raisers of medical funds randomly selected on the GoFundMe platform achieved their fundraising goals. Howard [14] pointed out that only a small proportion of medical fundraising campaigns are fully funded. The statistical data of 143,917 medical crowdfunding projects on Qschou.com show that only 7% of the projects successfully raised the target amount [15]. Therefore, a very important research goal is to identify the factors that affect the success rate of medical crowdfunding projects.

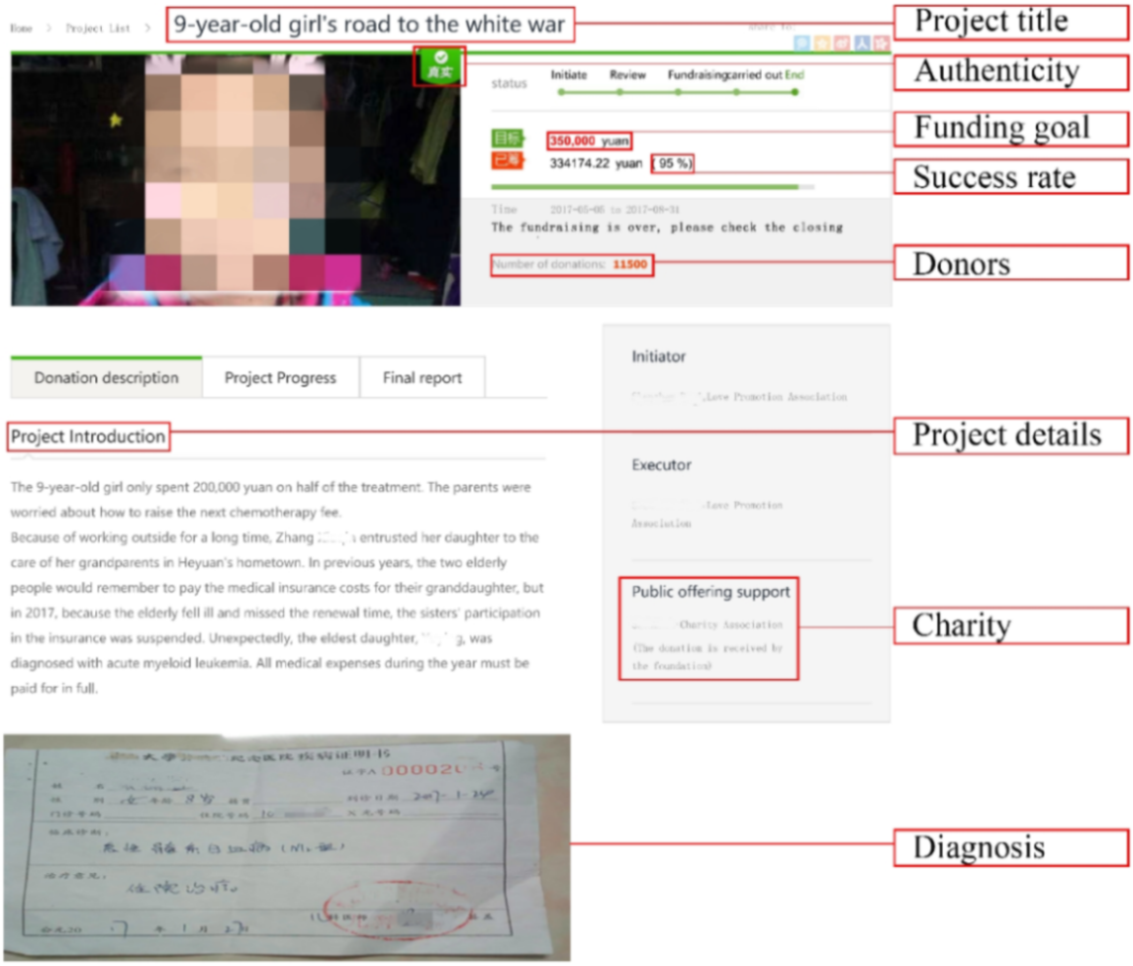
With the rapid development of medical crowdfunding, many studies have begun to examine this topic. Most researchers in this field have focused on the development of medical crowdfunding; the relationship with personal bankruptcy; and the fraud, privacy, and ethics issues that may be caused by medical crowdfunding [10,16-19]. However, little research has focused on the factors that impact the success of a medical crowdfunding project.

Medical crowdfunding is a new form of online crowdfunding. The role of the text on a fundraising page is important because project creators try to convince potential supporters with a compelling story [20]. Numerous studies have analyzed the effect of textual features on the success of online crowdfunding; however, determining whether these features also affect the success rate of medical crowdfunding projects is an extremely important and unsolved problem. Additionally, Aristotle's theory of persuasion has been widely used in the online crowdfunding literature [20-22]. For example, Majumdar and Bose [20] studied the impact of narrative text on the possibility of pizza donation according to this theory. However, there is no literature related to the influence of textual features on the success of medical crowdfunding based on this theory at present. Therefore, we proposed research question 1: According to previous research and Aristotle's theory of persuasion, do textual features have an important impact on the success of medical crowdfunding?

In medical crowdfunding, the project information generally consists of two parts: the project title and the project details, as shown in Figure 1. The project title is used to provide a brief overview of the crowdfunding project information using concise phrasing, and the project details are used to provide more complete information about the patient. Some studies have already investigated the impact of textual features on medical crowdfunding success; however, they focused on the impact of the text features of the project details and ignored the text features of the project title [23,24]. Medical crowdfunding projects are usually spread via social media in the form of links; therefore, potential donors often see the project title first. Accordingly, the textual features of the project title may affect the behavior of potential donors. If potential donors want to know more after reading the project title, they will further read the content of the project details. At this time, the text features of the project details may affect readers' donation behavior. Therefore, it is possible that the text features of both the project title and project details will affect donors' decision-making behavior, which in turn will affect the success of medical crowdfunding projects. Therefore, we proposed the following two research questions to explore the characteristics of the project title and project details that affect the success rate of medical crowdfunding projects:

- Research question 2: In the project title, what features can influence the success of medical crowdfunding?
- Research question 3: In the project details, what features can influence the success of medical crowdfunding?

Figure 1. Screenshot of a medical crowdfunding project.



The 9-year-old girl only spent 200,000 yuan on half of the treatment. The parents were worried about how to raise the next chemotherapy fee.

Because of working outside for a long time, Zhang [redacted] entrusted her daughter to the care of her grandparents in Heyuan's hometown. In previous years, the two elderly people would remember to pay the medical insurance costs for their granddaughter, but in 2017, because the elderly fell ill and missed the renewal time, the sisters' participation in the insurance was suspended. Unexpectedly, the eldest daughter, [redacted], was diagnosed with acute myeloid leukemia. All medical expenses during the year must be paid for in full.

Literature Review

Medical Crowdfunding

Medical crowdfunding is a form of donation-based online crowdfunding. Online crowdfunding involves enterprises or individuals obtaining, through the internet, financial support from a large group of participants who provide small amounts of money [25]. There are many types of crowdfunding such as reward-based crowdfunding, equity-based crowdfunding, debt-based crowdfunding, and donation-based crowdfunding [26]. Reward-based crowdfunding involves individuals contributing comparatively small amounts of money to projects in return for some kind of reward. Rewards can range from a simple token such as a thank-you postcard to a production version of the crowdfunded product [27]. Equity-based crowdfunding is a mechanism that enables broad groups of investors to fund startup companies and small businesses in return for equity [28]. Debt-based crowdfunding is based on the relationship between a debtor and lender, and possesses a certain yield rate [29]. Donation-based crowdfunding, also called charitable crowdfunding, requires donors to devote themselves to the public good with no expectation of obtaining material rewards [30].

Medical crowdfunding aims to help those who are unable to pay their medical bills by enabling them to receive donations from others, and its text narrative usually involves many personal characteristics of the requester. In recent years, medical crowdfunding has attracted extensive attention within the academic community. Specifically, Snyder et al [18] analyzed 80 medical crowdfunding activities as examples of how Canadians prove to others that they should finance the medical expenses of requesters, and these reasons were evaluated according to relevant theories of ethics. Renwick and Mossialos [31] examined possible economic risks and benefits in the development of medical crowdfunding. Gonzales et al [16] demonstrated that medical crowdfunders often balance two aspects of a fundraising campaign: assistance from others and perceived privacy risks. Zenone and Snyder [19] summarized four possible types of fraud in medical crowdfunding: faking/exaggerating illnesses, impersonation, and misapplication of funds. Moore [17] studied an ethical issue involving the role of donors in medical crowdfunding. The research of Burtch and Chan [10] indicated that the number of individual bankruptcy filings in the United States is experiencing a trend of gradual decrease as a result of the rapid advancement of medical crowdfunding. Bassani et al [32] discussed the formation process and development direction of medical crowdfunding platforms, explored the relationship between medical crowdfunding and the national health system, and found that substitution effects function when the coverage rate of public health is low. After analyzing news reports on medical crowdfunding in the United States and Canada, Murdoch et al [33] discovered that there is little reportage of the negative descriptions of crowdfunding campaigns (4.76%), and found that the patients' conditions were mostly cancer (49.11%) and rare diseases (36.01%).

In summary, previous studies have not directly elucidated the influencing factors of medical crowdfunding results but have rather probed the development of medical crowdfunding, and its impact on social and economic activities.

Influence of Textual Features on the Success of Online Crowdfunding

Crowdfunding projects are usually associated with a description of how the funds will be used [20]. In recent years, many studies have analyzed the impact of textual features on the success of online crowdfunding. Greenberg et al [34] found that some project information characteristics such as the project goal, project category, magnitude of the reward, project duration, whether the project is connected to social networks, and the number of words in the narrative can predict the success or failure of a reward-based crowdfunding campaign. Lukkarinen et al [28] discovered that the investment decisions of participants in equity-based crowdfunding are associated with the comprehensibility of the campaign and other easily grasped informational features to a great extent, with the traditional investment decision criteria used by venture capitalists or business angels having less importance. Lin [35] suggested that the language characteristics of the campaign requirements description, including readability, positivity, and deception cues, can predict the performance of the loan project.

Unlike other types of online crowdfunding, donation-based crowdfunding does not require financial returns, and the narrative usually contains little information about the economic benefits. Some scholars have explored the influence of textual features on the success of donation-based crowdfunding. Kshetri [36] pointed out that the identity of the fundraiser displayed in a donation-based crowdfunding campaign has a significant influence on donors' behavior. Xu [37] examined the impact of three manifestations (video, pictures, and text) on the success rate of donation-based crowdfunding. Majumdar and Bose [20] analyzed 5671 real data items and found that donors' behavior is influenced by many variables, including money-related words, the extent of use of negative emotions, and evidence relating to women and the authenticity of the information they provided. Additionally, Zubrickas [38] and Cason and Zubrickas [39] discussed how the inclusion of a refunding mechanism in the information can affect donors' decision-making behavior. Sasaki [40] explained the effect of the amount of other people's donations on subsequent behavior, and the empirical results showed that subsequent donors are more likely to increase their donations when the amount of the previous five donations increases. Cheng et al [41] proposed an optimal threshold in the number of donations provided by donation-based crowdfunding campaigns that promotes more donation activities.

Medical crowdfunding is a new form of donation-based crowdfunding. However, it is unknown whether the text characteristics of these projects affect their success rate. Only a few researchers have aimed to study this problem to date. Durand et al [23] indicated that medical crowdfunding for transplant surgery was more successful when the campaigns had longer descriptions, higher goal amounts, more positive emotions, and third-person narratives. Majumdar and Bose [20] discovered that reasonable and credible requests during a campaign can improve the chance of receiving donations. Xu and Wang [24] considered that certain messaging aspects such as constructing an image of being vulnerable and worthy of help, utilizing a tragic narrative with elements of Chinese traditional culture, and contrasting patients' experiences before and after the illness can arouse the sympathy of potential donors and prompt them to donate to patients. However, these studies analyzed the impact of only a small number of textual features on medical crowdfunding from a specific theoretical perspective. Whether the large number of textual features found in the online crowdfunding literature has an impact on the success rate of medical crowdfunding remains an unresolved research question. Moreover, these previous studies focused only on the influence of the textual features of the project details and ignored the role of the textual features of the project title. However, the project title also embodies a wealth of information that may play a prominent role in the success rate of medical crowdfunding projects.

Theory of Persuasion and Online Crowdfunding

As an important channel for people to understand the needs of a medical crowdfunding project, the main expectation of text narration is to persuade readers to donate to the requester. The type of persuasive strategy that can be used to improve the persuasiveness of the text is particularly significant. Aristotle supposed that rhetoric signifies "the faculty of observing in any given case the available means of persuasion" and proposed the theory of persuasion. This theory mainly includes three important methods of persuasion: Ethos, Logos, and Pathos [42]. Ethos means that speakers should show some authentic and reliable content to gain psychological recognition and the trust of the audience. Logos means that speakers should use logical statements to convince others. Pathos means that speakers should display their emotional factors and arouse the emotions of the audience [42,43].

Aristotle's theory of persuasion has also been applied to the field of online crowdfunding. Tirdatov [21] examined the 13 most-funded crowdfunding texts from Kickstarter, and found that all of them contained the three types of rhetorical styles mentioned in the theory of persuasion (ie, Ethos, Pathos, and Logos). Majumdar and Bose [20] studied the content factors of rationality, emotion, and credibility in text narratives, and found that the presence of rational and credible content in charity messages is positively related to the probability of receiving pizza donations, but this was not the case for the emotional content. Wang and Wu [22] explored the impact of photographic narratives on online medical crowdfunding campaigns. They found that Pathos (emotional-related) and Ethos (credible-related) photos have positive persuasion effects on medical crowdfunding performance, whereas Logos (rational-related) photos have a negative persuasion effect. However, the existing literature has not yet used the theory of persuasion to study the influence of persuasive strategies of textual features on the results of medical crowdfunding.

Complementing this line of existing research, this study empirically examined the effect of the textual features of medical crowdfunding projects on their success rate, taking 4903 real projects collected from the Tencent Charity platform in China as the research sample. Based on the theory of persuasion, we divided the appeals in the text into three categories (rational, emotional, and credibility appeals), explored them from the two dimensions of project title and project details, and analyzed them from two perspectives of existence and extent.

Methods

Study Design

To meet the research objectives, a research framework was developed. The process was divided into five main sections: (1) collecting and cleaning the dataset, (2) selecting the variables and building the research model, (3) extracting the variables, (4) performing econometric analysis, and (5) evaluating the results of the econometric analysis to extract knowledge.

Collecting and Cleaning the Dataset

There are two types of medical crowdfunding in China: public fundraisers and individual appeals. The "Charity Law" promulgated in 2016 stipulates that for public fundraisers, patients or their families must post fundraising information on the crowdfunding platform prescribed by the Civil Affairs Department of the State Council with the help of philanthropic organizations. These projects are publicly available on the crowdfunding platform where information is easily obtained. Individual appeals are published and disseminated exclusively on social media such as microblogs and cannot be publicly displayed on crowdfunding platforms. Compared to individual appeals, the information on crowdfunding platforms is more complete and has higher credibility, and the platforms have certain specifications for fundraisers so that the process is more normalized. Hence, the datasets in this study were obtained mainly from public fundraisers.

Our data sample was obtained from the Tencent Charity platform, which is one of the largest donation-based crowdfunding platforms in China and provides easy access to information. Tencent Charity was initiated by the Tencent Company in 2007. As of January 2019, its cumulative fundraising amount had exceeded US $700 million. We collected data using a crawler program, which we developed based on the Python language. The program first uses a Python third-party library ("requests") to obtain webpage data containing crowdfunding projects. It then uses regular expressions to parse and extract the required information, and finally saves the data in an Excel table. The pseudocode of the crawler program is provided in Multimedia Appendix 1.

We collected all disease crowdfunding projects displayed on Tencent Charity in January 2019. An initial total of 5400 projects were obtained. We deleted projects that were raising funds for certain groups through manual filtering, because the focus of this study was personal medical crowdfunding. The main process was as follows. First, two graduate students marked projects for which the donation objects were not individuals by reading crowdfunding text from the dataset. Second, a third student compared the results of the previous two students. If there was substantial divergence, the three students discussed it until they reached agreement. In the end, a final dataset containing information on 4903 medical crowdfunding projects containing both nontextual features (funding goal, fundraising amount, and donors) and textual features (project title and project details) was obtained. In the dataset, the average values of funding goal, fundraising amount, and number of donors were 149,673, 34,569, and 1467, respectively. We used the fundraising amount divided by the funding goal to calculate the project's fundraising success rate. Overall, the success rate of medical crowdfunding projects was found to be relatively low, with an average success rate of 0.31.

Selecting Variables and Building the Research Model

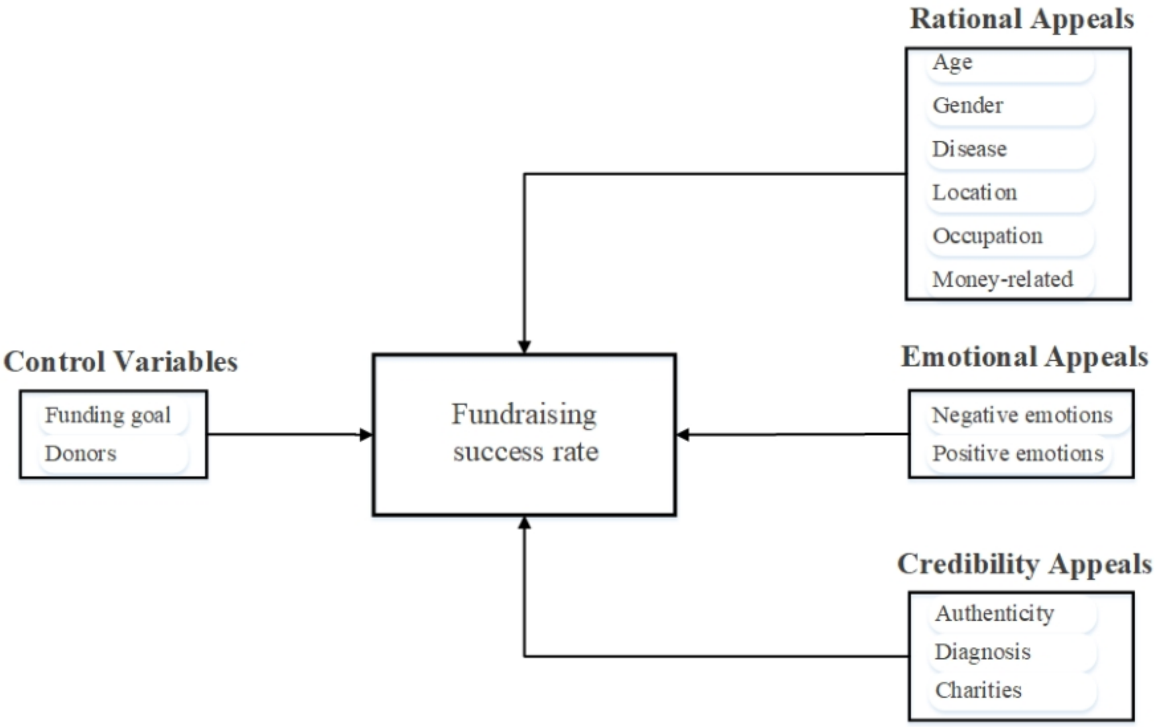
We first selected variables according to the previous literature on online crowdfunding, which showed that certain textual features play an important role in the success rate of online crowdfunding projects, as shown in Table 1. The main textual features that have been found to have a significant impact include disease description, patient age, patient gender, regional economic development level, patient occupation, emotional and economic descriptions, vivid language, inclusive language, money-related language, title length, project information quality, description elaborateness, video, images, number of updates, prior experience, and duration. These features constituted the candidate feature set of the crowdfunding texts. We then determined the variables related to medical crowdfunding based on this feature set, including textual and nontextual features (Table 1).

Table 1. Textual features affecting the success rate of online crowdfunding projects.

Reference	Online crowdfunding type	Research variables
Xu and Wang [24]	Donation-based crowdfunding	Illness narrative and the patients' identity such as disease severity and financial plight, tragic narrative strategies, etc
Aleksina et al [44]	Donation-based crowdfunding	Disease characteristics, medical research characteristics, organizational characteristics, comments, length, platform, etc
Kusumarani and Zo [45]	Donation-based crowdfunding	Perceived financial resources, perceived information resources, political interest, negative perception toward political situation, online community engagement, etc
Liang et al [46]	Reward-based crowdfunding	Funders' trust, project information quality, fundraiser's ability, project type and funding level, gender, age, education, income, prior experiences, website familiarity, etc
Petitjean [47]	Reward-based crowdfunding	Goal, geography, photo, video, website, updates, comments, Facebook friends, shares, etc
Durand et al [23]	Donation-based crowdfunding	Disease category, narrative perspective, gender, patient age group, location, goal amount, cumulative sentiment, sentiment range, description length
Majumdar et al [20]	Donation-based crowdfunding	Title length, reciprocity phrases, gratitude, popularity, comments received, account age, past participation, authentic, text length, presence of image, presence of monetary evidence, presence of negative emotions, presence of female references, extent of monetary evidence, extent of negative emotions, extent of female references
Kim et al [48]	Reward-based crowdfunding	Identity disclosure, prior experience, number of comments, number of updates, description elaborateness, campaign duration, funding goal, country, category
Allison et al [49]	Reward-based crowdfunding	Log funding, requested funding, adopted group, identity, positive narrative tone, etc
Bi et al [50]	Reward-based crowdfunding	Goal, duration, introduction word count, video count, "like" count, number of reviews
Lukkarinen et al [28]	Equity-based crowdfunding	Number of investors, amount raised, team rating, markets rating, concept rating, etc
Gorbatai et al [51]	Reward-based crowdfunding	Money-related language, vivid language, positive emotion, inclusive language, campaign duration, video, images, length of the campaign text, campaign goal, etc
This study	Donation-based crowdfunding	Age, gender, disease, location, occupation, money-related words, negative emotions, positive emotions, authenticity, diagnosis, charities, length of the project title and project details, funding goal, donors

We further built the research model based on Aristotle's theory of persuasion, which includes three important methods of persuasion: Ethos, Logos, and Pathos [42]. Some previous studies have applied the theory of persuasion to the study of human behavior [20,22,43]. Majumdar and Bose [20] categorized content factors of online crowding into rational, emotional, and credibility based on this theory. Similarly, in this study, we divided the persuasion strategy of medical crowdfunding text into three categories: rational, emotional, and credibility appeals, corresponding to Logos, Pathos, and Ethos, respectively. Age, gender, disease, location, occupation, and money-related words all describe actual information about the patient, highlighting the characteristics of the donation objects, and therefore constitute rational appeals. Negative emotions and positive emotions express the feelings of the seeker and focus on appealing to the sympathy of potential donors, and therefore belong to emotional appeals. Authenticity, diagnosis, and charities are used to prove that the crowdfunding project is true and reliable, and are therefore classified as credibility appeals. Moreover, consistent with previous research [50], the nontextual features of funding goal and donors are regarded as control variables. The research model is schematically depicted in Figure 2.

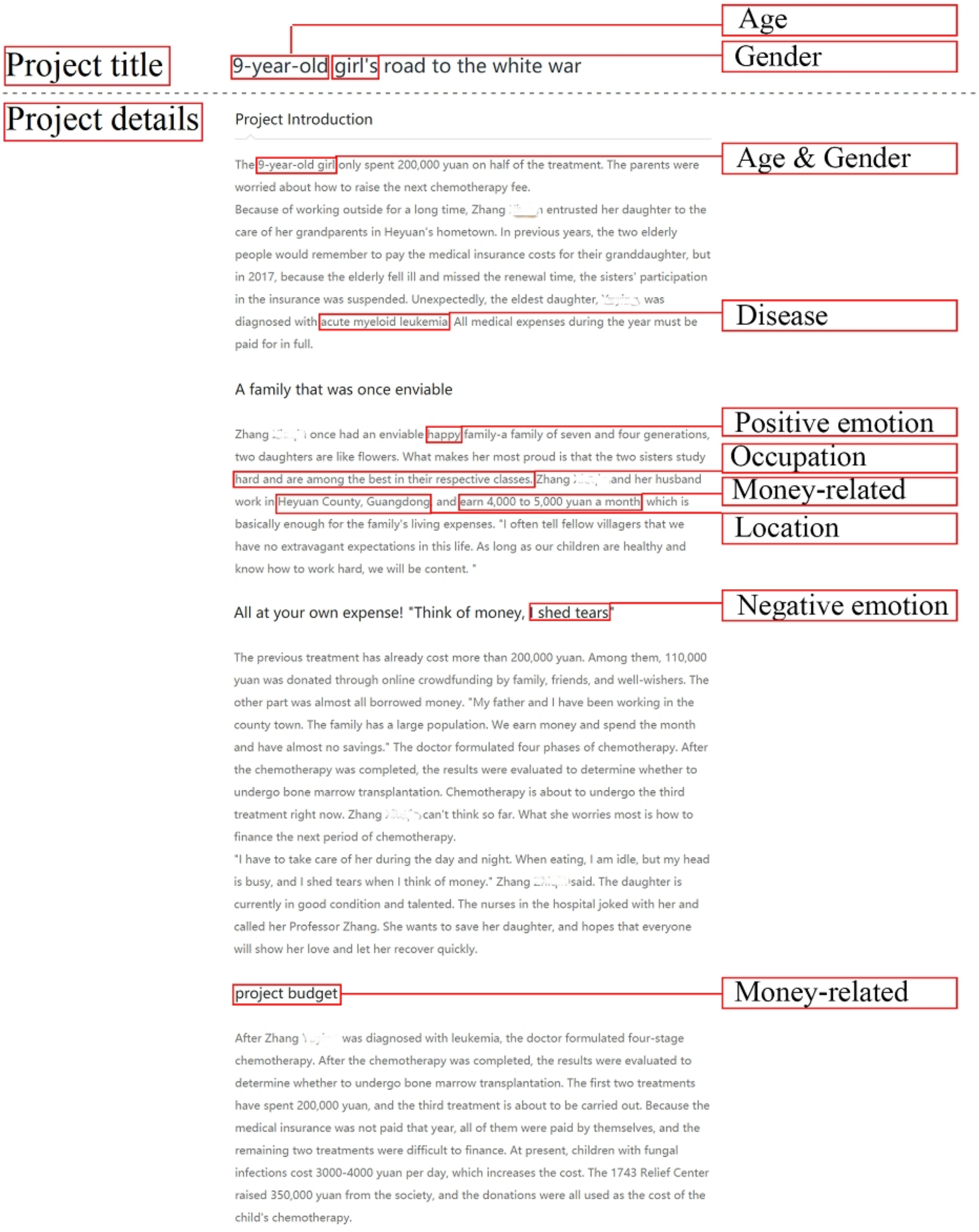
Figure 2. Research model.



Extracting Variables

In medical crowdfunding, some textual features appear in the project title, while others appear in the project details, and some may be reflected in both. Figure 3 shows a medical crowdfunding project from Tencent Charity with the textual features marked. We classified these variables into project title and project details. Moreover, a textual feature has two dimensions: whether the feature exists and the extent of the feature if it does exist. Majumdar and Bose [20] found that the presence of textual features (money-related terms and female references) and the extent of textual features (money-related terms, negative emotions, and female references) impact the success of donation-based crowdfunding. Therefore, we adopted a similar method to quantify all textual features in the dataset as existence variables or extent variables. Table 2 lists all of the variables considered with their descriptions.

Figure 3. Textual features of a medical crowdfunding project.



9-year-old girl's road to the white war

Project Introduction

The 9-year-old girl only spent 200,000 yuan on half of the treatment. The parents were worried about how to raise the next chemotherapy fee.

Because of working outside for a long time, Zhang [redacted] entrusted her daughter to the care of her grandparents in Heyuan's hometown. In previous years, the two elderly people would remember to pay the medical insurance costs for their granddaughter, but in 2017, because the elderly fell ill and missed the renewal time, the sisters' participation in the insurance was suspended. Unexpectedly, the eldest daughter, [redacted] was diagnosed with acute myeloid leukemia. All medical expenses during the year must be paid for in full.

A family that was once enviable

Zhang [redacted] once had an enviable happy family-a family of seven and four generations, two daughters are like flowers. What makes her most proud is that the two sisters study hard and are among the best in their respective classes. Zhang [redacted] and her husband work in Heyuan County, Guangdong and earn 4,000 to 5,000 yuan a month, which is basically enough for the family's living expenses. "I often tell fellow villagers that we have no extravagant expectations in this life. As long as our children are healthy and know how to work hard, we will be content. "

All at your own expense! "Think of money, I shed tears"

The previous treatment has already cost more than 200,000 yuan. Among them, 110,000 yuan was donated through online crowdfunding by family, friends, and well-wishers. The other part was almost all borrowed money. "My father and I have been working in the county town. The family has a large population. We earn money and spend the month and have almost no savings." The doctor formulated four phases of chemotherapy. After the chemotherapy was completed, the results were evaluated to determine whether to undergo bone marrow transplantation. Chemotherapy is about to undergo the third treatment right now. Zhang [redacted] can't think so far. What she worries most is how to finance the next period of chemotherapy.

"I have to take care of her during the day and night. When eating, I am idle, but my head is busy, and I shed tears when I think of money." Zhang [redacted] said. The daughter is currently in good condition and talented. The nurses in the hospital joked with her and called her Professor Zhang. She wants to save her daughter, and hopes that everyone will show her love and let her recover quickly.

project budget

After Zhang [redacted] was diagnosed with leukemia, the doctor formulated four-stage chemotherapy. After the chemotherapy was completed, the results were evaluated to determine whether to undergo bone marrow transplantation. The first two treatments have spent 200,000 yuan, and the third treatment is about to be carried out. Because the medical insurance was not paid that year, all of them were paid by themselves, and the remaining two treatments were difficult to finance. At present, children with fungal infections cost 3000-4000 yuan per day, which increases the cost. The 1743 Relief Center raised 350,000 yuan from the society, and the donations were all used as the cost of the child's chemotherapy.

Table 2. Description of the variables.

Variables	Description and measure
Success rate	Ratio of raised amount to target amount
Funding goal	The total amount to be raised by a fundraiser in a project (10,000)
Donors	Number of people contributing to a project
TPAge	Dummy variable; 1 if there is an age in the title, otherwise 0
TPGender	Dummy variable; 1 if there is a gender in the title, otherwise 0
TPDisease	Dummy variable; 1 if there is a disease description in the title, otherwise 0
TPOccupation	Dummy variable; 1 if there is an occupation in the title, otherwise 0
TPLocation	Dummy variable; 1 if there is a location in the title, otherwise 0
TPMoney	Dummy variable; 1 if there is a funding target in the title, otherwise 0
TPNWords	Dummy variable; 1 if there are negative words in the title, otherwise 0
TPPWords	Dummy variable; 1 if there are positive words in the title, otherwise 0
DPAge	Dummy variable; 1 if there is an age in the details, otherwise 0
DPGender	Dummy variable; 1 if there is a gender in the details, otherwise 0
DPDisease	Dummy variable; 1 if there is a disease description in the details, otherwise 0
DPOccupation	Dummy variable; 1 if there is an occupation in the details, otherwise 0
DPLocation	Dummy variable; 1 if there is a location in the details, otherwise 0
DPMoney	Dummy variable; 1 if there is a financial description in the details, otherwise 0
DPNWords	Dummy variable; 1 if there are negative words in the details, otherwise 0
DPPWords	Dummy variable; 1 if there are positive words in the details, otherwise 0
TLength	Number of words in the project title
TAge	Patient’s age in the project title
TGender	Patient’s gender in the project title
TDisease	Patient’s disease in the project title; 1 for mild disease and 0 for severe disease
TLocation	Patient’s location in the project title; sorted according to per capita GDP in 2018 of each city in descending order, and assign 1, 2, 3..., as shown in Multimedia Appendix 2
TOccupation	Patient’s occupation in the project title; classify patient’s occupation into nine categories and assign them values from 1 to 9, as shown in Multimedia Appendix 3
TNMWords	Number of mentions of money-related words in the project title
TNNWords	Number of negative emotion words in the project title
TNPWords	Number of positive emotion words in the project title
DLength	Number of words in the project details
DAge	Patient’s age in the project details
DGender	Patient’s gender in the project details
DDisease	Patient’s disease in the project details; 1 for mild disease and 0 for severe disease
DLocation	Patient’s location in the project details; sorted according to per capita GDP in 2018 of each city in descending order, and assign 1, 2, 3..., as shown in Multimedia Appendix 2
DOccupation	Patient’s occupation in the project details; classify patient’s occupation into nine categories and assign them values from 1 to 9, as shown in Multimedia Appendix 3
DNMWords	Number of mentions of money-related words in the project details
DNNWords	Number of negative emotion words in the project details
DNPWords	Number of positive emotion words in the project details
Authenticity	1 indicates that the project has been verified by the platform; otherwise, 0
Diagnosis	1 indicates that the project includes a hospital diagnosis certificate; otherwise, 0

Variables	Description and measure
Charities	1 indicates that charities are involved in the project; otherwise, 0

Automatic extraction and manual filtering were used to extract the textual features and corresponding variables. The patient's age, gender, disease, occupation, and location, as well as money-related items were manually extracted because it was difficult to automatically process these variables, as the writing style of the text was personal and had no fixed or standardized format. For each feature, we determined the value of its existence variable and extent variable in the project title and project details. For example, for patient gender, we determined the values of four variables: *TPGender*, *DPGender*, *TGender*, and *DGender*. The manual extraction process took approximately 2 months to complete and was divided into two main stages. First, two graduate students independently extracted each feature by reading each individual text from the dataset. Second, a third student compared the results of the previous two students. If there was substantial divergence, the three students discussed it until they reached agreement. However, owing to the requirements of the Tencent Charity platform, all projects meet the credibility appeals. All of the crowdfunding projects we collected were certified by the platform, with hospital diagnostic reports and charities participation. Therefore, authenticity, diagnosis, and charities were set to a value of 1.

Other variables, including *TLength*, *DLength*, *TPNWords*, *TPPWords*, *TNNWords*, *TNPWords*, *DPNWords*, *DPPWords*, *DNNWords*, and *DNPWords*, were automatically extracted using Python and the HowNet lexicon, which is the Chinese version of WordNet and has been widely used for text mining and sentiment analysis [52-54]. First, we calculated the number of words in the project title and project details to obtain the values of the variables *TLength* and *DLength*. Next, we used a Chinese word segmentation component called Jieba in Python to segment the project title and project details, and then compared the words in the project title with the positive emotional vocabulary in the HowNet dictionary. If the match was successful, the value of the variable *TPPWords* was recorded as 1 and the value of the variable *TNPWords* was increased by 1. The values of other variables were calculated in the same manner.

Building the Econometric Model

To better appreciate the effects of the textual features of the project title and project details, we used multiple regression analysis to construct five econometric models. Model 1 was the control-only model. Model 2 considered the presence variables of the project title and the control variables. Model 3 added the presence variables of the project details on the basis of Model 2. Model 4 considered the extent variables of the project title and the control variables. Model 5 added the extent variables of the project details on the basis of Model 4. Since the values of authenticity, diagnosis, and charities were all constant, indicating that there was no difference in the credibility appeals among the crowdfunding projects, these three variables were removed when building the econometric models.

- Model 1:

  $\text{Success rate} = \alpha_0 + \alpha_1\text{Funding goal} + \alpha_2\text{Donors} + \varepsilon$
- Model 2:

  $\text{Success rate} = \alpha_0 + \alpha_1\text{Funding goal} + \alpha_2\text{Donors} + \alpha_3\text{TPAge} + \alpha_4\text{TPGender} + \alpha_5\text{TPDisease} + \alpha_6\text{TPOccuption} + \alpha_7\text{TPLocation} + \alpha_8\text{TPMoney} + \alpha_9\text{TPWords} + \alpha_{10}\text{TPPWords} + \varepsilon$
- Model 3:

  $\text{Success rate} = \alpha_0 + \alpha_1\text{Funding goal} + \alpha_2\text{Donors} + \alpha_3\text{TPAge} + \alpha_4\text{TPGender} + \alpha_5\text{TPDisease} + \alpha_6\text{TPOccuption} + \alpha_7\text{TPLocation} + \alpha_8\text{TPMoney} + \alpha_9\text{TPWords} + \alpha_{10}\text{TPPWords} + \alpha_{11}\text{DPAge} + \alpha_{12}\text{DPGender} + \alpha_{13}\text{DPDisease} + \alpha_{14}\text{DPOccupation} + \alpha_{15}\text{DPLocation} + \alpha_{16}\text{DPMoney} + \alpha_{17}\text{DPNWords} + \alpha_{18}\text{DPPWords} + \varepsilon$
- Model 4:

  $\text{Success rate} = \alpha_0 + \alpha_1\text{Funding goal} + \alpha_2\text{Donors} + \alpha_3\text{TLength} + \alpha_4\text{TAge} + \alpha_5\text{TGender} + \alpha_6\text{TDisease} + \alpha_7\text{TOccupation} + \alpha_8\text{TLocation} + \alpha_9\text{TNMWords} + \alpha_{10}\text{TNNWords} + \alpha_{11}\text{TNPWords} + \varepsilon$
- Model 5:

  $\text{Success rate} = \alpha_0 + \alpha_1\text{Funding goal} + \alpha_2\text{Donors} + \alpha_3\text{TLength} + \alpha_4\text{TAge} + \alpha_5\text{TGender} + \alpha_6\text{TDisease} + \alpha_7\text{TOccupation} + \alpha_8\text{TLocation} + \alpha_9\text{TNMWords} + \alpha_{10}\text{TNNWords} + \alpha_{11}\text{TNPWords} + \alpha_{12}\text{DLength} + \alpha_{13}\text{DAge} + \alpha_{14}\text{DGender} + \alpha_{15}\text{DDisease} + \alpha_{16}\text{DOccupation} + \alpha_{17}\text{DLocation} + \alpha_{18}\text{DNMWords} + \alpha_{19}\text{DNNWords} + \alpha_{20}\text{DNPWords} + \varepsilon$

Results

Descriptive Statistics

Table 3 provides the descriptive statistics of the variables. *DPGender*, *DPDisease*, *Authenticity*, *Diagnosis*, and *Charities* were constant, and were therefore removed from subsequent analyses. The correlations between the variables are provided in Multimedia Appendix 4 and Multimedia Appendix 5. Given the overall lack of a strong correlation between the variables, we proceeded to the regression analysis.

Table 3. Descriptive statistics of variables.

Variables	Maximum	Minimum	Median	Mean	Variance
Success rate	1.47	0.00	0.12	0.31	0.14
Funding goal	350.00	1.00	10.00	14.97	241.11
Donors	77,760.00	5.00	311.00	1466.59	17,465,022.11
TPAge	1.00	0.00	0.00	0.14	0.12
TPGender	1.00	0.00	0.00	0.34	0.23
TPDescription	1.00	0.00	0.00	0.40	0.24
TPLocation	1.00	0.00	0.00	0.01	0.01
TPOccupation	1.00	0.00	0.00	0.10	0.09
TPMoney	1.00	0.00	0.00	0.005	0.005
TPNWords	1.00	0.00	0.00	0.03	0.03
TPPWords	1.00	0.00	0.00	0.19	0.15
DPAge	1.00	0.00	1.00	0.85	0.12
TPGender	1.00	1.00	1.00	1.00	0.00
TPDisease	1.00	1.00	1.00	1.00	0.00
DPLocation	1.00	0.00	1.00	0.84	0.13
DPOccupation	1.00	0.00	1.00	0.93	0.07
DPMoney	1.00	0.00	1.00	0.98	0.02
DPNWords	1.00	0.00	1.00	0.93	0.07
DPPWords	1.00	0.00	1.00	0.998	0.002
TLength	12.00	4.00	9.00	8.46	0.75
TAge	96.00	0.00	4.00	7.07	97.28
TGender	1.00	0.00	0.00	0.42	0.24
TDisease	1.00	0.00	0.00	0.02	0.02
TLocation	33.00	1.00	20.00	17.23	118.57
TOccupation	9.00	1.00	7.00	6.22	3.39
TNMWords	1.00	0.00	0.00	0.005	0.005
TNNWords	2.00	0.00	0.00	0.03	0.03
TNPWords	2.00	0.00	0.00	0.20	0.17
DLength	3785.00	121.00	1355.00	1370.20	218,283.22
DAge	96.00	0.00	8.00	13.15	196.38
DGender	1.00	0.00	1.00	0.60	0.24
DDisease	1.00	0.00	0.00	0.03	0.03
DLocation	43.00	1.00	18.00	18.29	74.85
DOccupation	9.00	1.00	7.00	6.73	2.11
DNMWords	22.00	0.00	3.00	3.87	6.58
DNNWords	30.00	0.00	4.00	4.35	12.11
DNPWords	68.00	0.00	16.00	16.96	59.48
Authenticity	1.00	1.00	1.00	1.00	0.00
Diagnosis	1.00	1.00	1.00	1.00	0.00
Charities	1.00	1.00	1.00	1.00	0.00

Regression Results

We first estimated the effect of control variables on the success rate of project fundraising with Model 1. As shown in Table 4, the funding goal and number of donors were significantly related to the success rate. Specifically, the funding goal had a negative impact on the success rate of fundraising, and the coefficient of donors (19.00, *P*<.001) significantly and positively affected the success rate.

Based on these control variables, we then increased the number of existence variables in the project title with Model 2. As shown in Table 4, the effects of the funding goal and number of donors were significant in this model, coinciding with the findings for Model 1. Moreover, in rational appeals, the presence of a patient's disease and occupation had a positive impact on the success rate of crowdfunding, whereas the presence of a patient's age and gender had a negative impact. The presence of location and monetary evidence exhibited no significant effect. In emotional appeals, the coefficient of *TPNWords* (–3.83, *P*=.04) indicated that the presence of negative emotions in the project title played a negative role in the fundraising success rate. The presence of positive emotions was insignificant, which shows that the persuasive strategy of positive emotions had no effect.

Additionally, we added the factors of project details to Model 2 to build Model 3, so as to estimate the role of project details. As indicated in Table 4, the control variables still exhibited significant effects, consistent with Models 1 and 2. The significant variables in the project title were also the same as those found for Model 2. In rational appeals of the project details, the coefficients of *DPOccupation* and *DPMoney* were 3.07 (*P*=.01) and 5.01 (*P*=.02), respectively, reflecting that the presence of the patient's occupation and money-related words had a positive relationship on the fundraising success rate. The presence of the patient's age had a negative relationship and the presence of the patient's location had no significant effect, which indicates that this information cannot persuade a donor. In emotional appeals, the presence of negative emotions was found to play a positive role in the success rate of fundraising, while the presence of positive emotions played a negative role, implying that the presence of negative emotions, rather than positive emotions, is more persuasive to donors.

Table 4. Regression results of existence variables.

Variables	Model 1 (adjusted R^2=0.637)			Model 2 (adjusted R^2=0.639)			Model 3 (adjusted R^2=0.643)		
	β	*P* value	VIF[a]	β	*P* value	VIF	β	*P* value	VIF
Control									
Funding goal	–20.59	<.001	1.076	–20.51	<.001	1.114	–20.53	<.001	1.179
Donors	19.00	<.001	1.076	19.01	<.001	1.083	18.90	<.001	1.097
Title									
TPAge	—[b]	—	—	–3.27	<.001	1.035	–2.96	<.001	1.051
TPGender	—	—	—	–2.29	.001	1.029	–2.29	.001	1.043
TPDisease	—	—	—	1.34	.04	1.032	1.41	.03	1.037
TPLocation	—	—	—	–4.90	.08	1.006	–5.45	.05	1.009
TPOccupation	—	—	—	2.67	.01	1.021	2.51	.02	1.025
TPMoney	—	—	—	–0.68	.88	1.006	–1.85	.68	1.011
TPNWords	—	—	—	–3.83	.04	1.008	–3.76	.04	1.008
TPPWords	—	—	—	–1.02	.21	1.015	–1.02	.21	1.017
Details									
DPAge	—	—	—	—	—	—	–4.36	<.001	1.039
DPLocation	—	—	—	—	—	—	–1.20	.17	1.031
DPOccupation	—	—	—	—	—	—	3.07	.01	1.041
DPMoney	—	—	—	—	—	—	5.01	.02	1.004
DPNWords	—	—	—	—	—	—	2.55	.04	1.042
DPPWords	—	—	—	—	—	—	–34.65	<.001	1.013

[a] VIF: variance inflation factor.

[b] —: not included in model.

Furthermore, we estimated the extent variables of the project title and project details to explore how the extent of a feature, if it exists, affects the fundraising results. Table 5 demonstrates the regression results for the extent variables of the project title and the control variables in Model 4. The funding goal exhibited a significant negative impact and the number of donors exhibited a significant positive effect, consistent with the results of the other models. None of the extent variables in the project title was statistically significant, indicating that the extent of the features in the project title causes no discernible effect on fundraising success rate.

Moreover, we examined the impact of the extent variables in the project details in Model 5. As shown in Table 5, the control variables also significantly affected the fundraising results. In addition, the length of the description in the project details exhibited a significant negative effect on fundraising success. *DNMWords* and *DNNWords* emerged as positive predictors of project success, which demonstrates that the more money-related and negative emotion words there are in the project details, the higher the success rate. The coefficient of *DAge* (–2.11, *P*<.001) suggests that younger patients are more likely to fundraise successfully in medical crowdfunding. The coefficient of *DOccupation* was also negative, showing that for patients whose occupations are national civil servant, professional skill worker, clerk, business and service worker, solider, child, student, and public-spirited person, the success rate of fundraising will decrease sequentially.

Table 5. Regression results of extent variables.

Variables	Model 4 (adjusted R^2=0.625)			Model 5 (adjusted R^2=0.613)		
	β	*P* value	VIF[a]	β	*P* value	VIF
Control						
Funding goal	–19.51	<.001	1.352	–17.76	<.001	1.796
Donors	19.88	<.001	1.185	18.46	<.001	1.297
Title[b]						
TLength	–12.29	.25	1.049	–4.73	.68	1.106
TAge	–1.22	.31	1.242	1.01	.48	1.553
TGender	–1.72	.29	1.056	–3.37	.23	1.243
TDisease	8.41	.53	1.009	14.09	.54	2.056
TLocation	–7.48	.07	1.005	–8.01	.05	1.010
TOccupation	.33	.36	1.051	.62	.11	1.080
TNNWords	–9.00	.47	1.039	–8.57	.49	1.064
TNPWords	1.49	.82	1.025	2.68	.69	1.030
Details						
DLength	—[c]	—	—	–11.73	.01	3.015
DAge	—	—	—	–2.11	<.001	1.010
DGender	—	—	—	.84	.67	1.192
DDisease	—	—	—	–.43	.98	2.097
DLocation	—	—	—	–.26	.74	1.115
DOccupation	—	—	—	–3.36	.004	1.254
DNMWords	—	—	—	5.78	.02	1.152
DNNWords	—	—	—	5.21	.02	1.585
DNPWords	—	—	—	–6.52	.08	2.568

[a] VIF: variance inflation factor.

[b] TNMWords was missing the correlation coefficient, and was therefore removed.

[c] —: not included in model.

Robustness Test

We used two methods to test the robustness of these empirical results: increasing sample data from other platforms and eliminating specific samples.

First, we increased the sample data to verify whether the results will be affected by platform differences. We collected 382 crowdfunding projects from the “Fun in Funding” platform [55], which is also a well-known medical crowdfunding platform in China. We use the new sample dataset to reexamine Model 3, and the results are presented in Table 6; the regression results using the original sample data (Table 4) are also shown for comparison. In the project title, except for the presence of the patient’s disease that became insignificant, the conclusions of the remaining variables were consistent. In the project details, the presence of the location changed from insignificant in the original model to significant with the added data, and there was no difference in the results for the other variables.

Second, specific samples were eliminated to test whether they had a major driving effect on the results. There was a relatively large number of samples with a small success rate in the original dataset. Therefore, we removed the sample data below the average success rate (0.31) and used the remaining sample data to perform multiple regression analysis on Model 3. Table 6 shows that in the project title, the presence of the patient's age and gender exhibited significant effects, coinciding with the previous conclusions. The presence of disease and location was significant at a 90% CI, which can also be statistically explained. The presence of occupation and negative emotions changed from significant to insignificant, whereas the presence of positive emotions transformed from insignificant to significant. By contrast, in the project details, the presence of money-related words became insignificant, and there was no change in the results for the other variables.

In summary, after using the two methods to complete the robustness test, we found that the results were mainly consistent with those obtained previously, indicating that our findings are relatively robust.

Table 6. Regression results of robustness tests.[a]

Variables	Model 1 (adjusted R^2=0.643)			Model 2 (adjusted R^2=0.590)			Model 3 (adjusted R^2=0.254)		
	β	*P* value	VIF[b]	β	*P* value	VIF	β	*P* value	VIF
Control									
Funding goal	–20.53	<.001	1.179	–19.74	<.001	1.167	–17.25	<.001	2.599
Donors	18.90	<.001	1.097	17.62	<.001	1.097	11.41	<.001	2.555
Title									
TPAge	–2.96	<.001	1.051	–2.11	.02	1.051	–3.76	.02	1.053
TPGender	–2.29	.001	1.043	–2.04	.003	1.038	–2.54	.04	1.057
TPDisease	1.41	.03	1.037	.90	.18	1.034	2.16	.07	1.054
TPLocation	–5.45	.05	1.009	–5.22	.07	1.008	–10.46	.03	1.021
TPOccupation	2.51	.02	1.025	2.34	.03	1.023	1.68	.33	1.031
TPMoney	–1.85	.68	1.011	–1.01	.83	1.011	2.93	.69	1.026
TPNWords	–3.76	.04	1.008	–4.84	.005	1.011	–1.24	.72	1.024
TPPWords	–1.02	.21	1.017	–.78	.34	1.018	–4.06	.007	1.038
Details									
DPAge	–4.36	<.001	1.039	–4.89	<.001	1.033	–4.55	.003	1.081
DPLocation	–1.20	.17	1.031	–2.06	.02	1.029	0.36	.80	1.041
DPOccupation	3.07	.01	1.041	3.18	.01	1.034	4.48	.03	1.082
DPMoney	5.01	.02	1.004	6.11	.005	1.003	2.43	.65	1.013
DPNWords	2.55	.04	1.042	2.65	.03	1.037	6.51	.001	1.058
DPPWords	–34.65	<.001	1.013	–36.57	<.001	1.012	–14.92	.05	1.025

[a] *P*<.10 at 90% CI is considered to be significant.

[b] VIF: variance inflation factor.

Discussion

Principal Findings

According to 4903 medical crowdfunding projects obtained from the Tencent Charity platform in China and using the theory of persuasion as the theoretical basis, this study applied multiple regression analysis to examine the impact of the existence variables and extent variables in the project title and project details on the success rate of medical crowdfunding. Owing to the requirements of the platform, all projects met the three aspects (authenticity, diagnosis, and charities) of credibility appeals; hence, we discuss the research results from two categories of appeals: rational and emotional appeals.

First, in rational appeals of the project title, the regression results of the existence variables showed that the presence of the patient's disease and occupation had a positive impact on the success rate of the medical crowdfunding project. Providing a narrative of the patient's disease and occupation allows potential donors to quickly understand who is raising money for what disease in a specific crowdfunding project. The presence of the patient's age and gender had a negative impact on the success rate, indicating that the seeker may not need to describe these attributes of the fundraiser. In emotional appeals of the project title, the presence of negative words had a negative impact on the success rate, which shows that people do not wish to see the requesters' negative emotions in the title description.

With respect to rational appeals of the project details, analysis of the existence variables suggested that the presence of a patient's occupation and money-related words is positively correlated with the fundraising success rate. The patient's occupation can arouse readers' pity, and money-related words can accurately describe the economic plight of the requesters. Majumdar and Bose [20] also suggested that the presence of monetary evidence can promote the success of crowdfunding. The presence of the patient's age was negatively correlated with the success rate, which indicates that the patient's age does not need to be described in the project details. In emotional appeals of the project details, the presence of negative emotions exhibited a positive impact on the fundraising success rate. However, Majumdar and Bose [20] suggested that the presence of negative emotions had no significant effect because almost all requests describe the suffering of the seeker. We consider this difference to be related to differences in the primary concerns of people in different countries. Chinese donors intend to understand the distress that patients are suffering, which can only be expressed accurately with negative phrases. We found that the presence of positive emotions exhibited a negative effect on success rate, whereas Durand et al [23] proposed that positive emotions contributed to the success of organ transplantation crowdfunding projects. The reason for this difference may be that the physically disabled but optimistic image in crowdfunding for organ transplantation is more persuasive, which is not the case with crowdfunding for medical diseases.

Furthermore, we estimated the extent variables of the project's textual features. The results reflect that in the project details, the patient's age, occupation, and text length had a negative impact on the fundraising success rate, whereas the numbers of money-related words and negative emotions words had a positive impact. We further analyzed the causes of these results in detail. First, we suggest that the number of money-related words has a significant positive impact on the success rate of fundraising, which coincides with the findings of Majumdar and Bose [20] and Xu and Wang [24]. Second, we found that the number of negative phrases has a positive impact on the success rate, whereas Majumdar and Bose [20] came to the opposite conclusion. They believed that the excessive use of negative appeals may seem manipulative to the potential donor. The reason for this difference may be related to the different cultures of multiple countries, as Chinese donors may be more likely to be moved by descriptions of patients' suffering that use more negative phrases. Xu and Wang [24] suggested that using a tragic narrative strategy can arouse the sympathy of Chinese potential donors. Third, we found that the length of the detailed narrative negatively affects the success rate, which is also inconsistent with the results of Majumdar and Bose [20]. The reason for this conflict may be that a long description can make people impatient and cause them to lose interest in donating. Iyengar and McGuire [56] suggested that humans are "cognitive misers," meaning that they will predigest limited information and make a rapid decision according to the principle of least possible effort. Fourth, Liang et al [46] proposed that age has no significant impact on reward-based crowdfunding, whereas our results show that young patients are more likely to succeed in medical crowdfunding because they more easily arouse people's sympathy. Similar to this conclusion, Ren et al [57] found that people are generally more generous with fundraising projects targeting children than adults. Finally, our study suggests that in the project details narrative of medical crowdfunding, for patients whose occupations are national civil servant, professional skill worker, clerk, business and service worker, solider, child, student, and public-spirited person, the success rate of fundraising will decrease sequentially.

Our results have clear implications with respect to the impact of the funding goal and donors as control variables. These two variables had evident significant effects on the fundraising success rate. Specifically, the funding goal was negatively correlated with the success rate, which is in accordance with the conclusions of Gorbatai and Nelson [51] and Kim et al [48], whereas the number of donors had a positive impact on the success rate, coinciding with the results of Aleksina et al [44], which revealed that sharing or forwarding information about crowdfunding activities on social networks can result in more financial support and an increased fundraising success rate.

Theoretical Implications

This study makes the following four theoretical contributions. First, this paper enriches and expands research in the realm of medical crowdfunding. The success rate of medical crowdfunding worldwide is generally low at present, yet few previous studies have directly explored the impact of textual features on the success rate of medical crowdfunding projects. This study used real projects to study the roles of textual features of the project title and project details, and the findings can provide a reference for researchers in medical crowdfunding.

Second, this study supplements and improves the online crowdfunding literature. Existing studies have mainly discussed the effect of the project details features but ignored the project title. By examining the project title features, we found that the presence of a patient's disease and occupation has a significant positive impact on the success rate of fundraising, whereas the presence of age, gender, and negative words has a negative impact. Thus, the role of the project title cannot be overlooked.

Third, the results of this study also contribute to the literature on persuasion theory. Although there have been many studies on human behavior based on the theory of persuasion, this study is the first to use this theory to explore the effects of textual persuasion strategies in the field of medical crowdfunding. Specifically, there is very little research on the persuasive strategy of the project title in online crowdfunding, although we discovered that this also has a contribution to fundraising results. For rational appeals, the presence of the patient's age, gender, disease, and occupation in the project title exhibited significant effects on the success rate of fundraising, whereas the presence of monetary evidence exhibited no important effect. For emotional appeals, the presence of negative emotions in the project title played a negative role on the success rate, whereas the presence of positive emotions had an insignificant effect. Moreover, the persuasion strategy of the project details in medical crowdfunding also differs from that in other types of crowdfunding. Liang et al [46] proposed that age has no significant impact on reward-based crowdfunding, whereas our results showed that young patients are more likely to successfully raise funds. There is no literature proposing that occupation has an important impact, although our results indicated that the presence of the patient's occupation makes the project more persuasive. Consistent with the findings of Majumdar and Bose [20], we suggest that the presence of monetary evidence can better persuade people to donate, and this persuasion effect increases with a greater frequency of that presence. For emotional appeals, the persuasion strategy of medical crowdfunding also diverges from that of other crowdfunding types. Majumdar and Bose [20] indicated that the presence of negative emotions had no significant effect on obtaining pizza donations. Durand et al [23] discovered that positive emotions contribute to the success of organ transplantation crowdfunding projects. However, our results indicate that the presence of negative emotions and positive emotions significantly affect the success rate of crowdfunding, but that the former affects the rate positively while the latter affects it negatively.

Fourth, this paper is beneficial for the online donation literature. In recent years, many studies have explained the donation behaviors of online users through external factors such as the celebrity effect and individual factors [30,58,59]. Behl and Dutta [60] also found that gamification has a positive impact on donor behavior on crowdfunding platforms for disaster relief operations. This study delineates how textual features in medical crowdfunding affect users' online donation behaviors, showing that the presence of a patient's disease and occupation in the project title has a positive relationship with the fundraising success rate, whereas the presence of age, gender, and negative words has a negative relationship. Moreover, the presence of the patient's occupation, monetary evidence, and negative emotions in the project details plays a positive role in the success rate, whereas the presence of the patient's age and positive emotions plays a negative role. Therefore, these features of the crowdfunding text may also affect the donation behaviors of online users. The conclusions can therefore provide assistance for research in the field of online donations.

Practical Implications

On the basis of our findings and previous research, we propose three implications for fundraisers from two categories (rational and emotional appeals) of the theory of persuasion and the three perspectives of the project title, project details, and nontextual features. First, for rational appeals of the project title, the results show that the presence of the patient's disease and occupation is positively related to the success rate of the medical crowdfunding project; thus, fundraisers should be advised to clearly describe the patient's illness and occupation in the project title. We also found that the presence of patient's age and gender is negatively related to the success rate, implying that age and gender are unnecessary in the narrative of the project title. For emotional appeals of the project title, the presence of negative words has a negative impact on the success rate, which suggests that the requester does not need to include negative emotions in the description of the project title.

Second, for rational appeals of the project details, the results indicate that the presence of a patient's occupation and money-related words has a positive impact on the fundraising success rate, which suggests that the description can include information about the patient's occupation and finances. Projects that use more money-related words are more likely to succeed, and the success rate will gradually decrease when the patient's occupation is national civil servant, professional skill worker, clerk, business and service worker, solider, child, student, and public-spirited person. In addition, the presence of the patient's age has a negative impact on the success rate, implying that the patient's age can be omitted from the description of the project details. However, if the patient is young, this information should be stated in the project details. For emotional appeals of the project details, the presence of negative words has a positive impact on the fundraising success rate, while the presence of positive words has the opposite effect, indicating that using as many negative phrases as possible is conducive to promoting fundraising success. The last suggestion is that the description in the details should be concise and powerful, and should not be too long.

Third, our investigation of the control variables showed that the fundraising success rate increases with increases in the number of donors but decreases with increases in the funding goal. This finding implies that requesters need to set appropriate goals when raising funds, and should widely disseminate and share their projects on social networks to seek more donations instead of forwarding the projects only to relatives and friends.

Finally, our study also has important implications for crowdfunding platform managers. Specifically, they can use a method similar to that presented herein to explore factors that are of concern to users in other types of crowdfunding text narratives. The managers of medical crowdfunding platforms may take our conclusions as a reference and provide proposals for fundraisers who have no idea how to create the crowdfunding text. Additionally, we recommend that some tags be intercalated in crowdfunding projects to highlight critical information that will be beneficial to the user experience of both fundraisers and donors.

Limitations and Future Research

Although this study makes several contributions, there are still some limitations that need to be recognized. First, our work considered the effects of textual features on the success of medical crowdfunding and did not examine noncontent features other than the funding goal and donors, such as the number of updates. Second, the process of feature extraction was somewhat cumbersome and time-consuming, and should be improved in the future. For example, future research can consider using machine learning and artificial intelligence to improve the efficiency of feature extraction. Third, this study collected only data related to the crowdfunding projects and excluded donor information. Future research can include donor information data and further analyze the impact of textual features on different donors. Fourth, this study only used fundraising projects from a Chinese crowdfunding platform, and therefore it is unclear whether the conclusions will be applicable to crowdfunding platforms from other countries. Thus, future research should be extended to collect data from multiple crowdfunding platforms in different countries for exploring the differences between crowdfunding in China and other countries from a cultural perspective. Finally, this study did not compare the effect of credibility appeals because of the limitation of the dataset.

Conclusion

This study focused on the effect of the textual features of medical crowdfunding projects on their success rate, taking 4903 real projects collected from the Tencent Charity platform in China as the research sample. We developed our research model according to previous studies and Aristotle's theory of persuasion. We further divided the textual features of the project title and project details into three categories of rational, emotional, and credibility appeals, which were analyzed from two perspectives of existence and extent. We show how some of the persuasive appeals and nontextual cues have a significant impact on fundraising success rates. The results show that in addition to the project details, the features of the project title also have an important impact on the success rate of fundraising. These findings provide important theoretical and managerial implications for medical crowdfunding. Additionally, our research will be beneficial to all donation-based charity platforms and requesters with medical fundraising needs.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (grant numbers 71764006, 71861013, 71861014, and 71762017), Hainan Provincial Natural Science Foundation of China (grant number 720RC572), Science Research Foundation of Hainan University (grant number kyqd(sk)1932), and the Educational Commission of Jiangxi Province of China (grant number GJJ200503).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The pseudocode of the crawler program.
[PDF File (Adobe PDF File), 151 KB - jmir_v23i6e22395_app1.pdf]

Multimedia Appendix 2
Quantification of location variables.
[PDF File (Adobe PDF File), 24 KB - jmir_v23i6e22395_app2.pdf]

Multimedia Appendix 3
Quantification of occupation variables.
[PDF File (Adobe PDF File), 181 KB - jmir_v23i6e22395_app3.pdf]

Multimedia Appendix 4
Correlation of existence variables.
[PDF File (Adobe PDF File), 52 KB - jmir_v23i6e22395_app4.pdf]

Multimedia Appendix 5
Correlation of extent variables.
[PDF File (Adobe PDF File), 68 KB - jmir_v23i6e22395_app5.pdf]

References

1. 2019 Global health care outlook: Shaping the future. Deloitte. URL: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-hc-outlook-2019.pdf [accessed 2020-06-20]
2. World Health Statistics 2018: Monitoring health for the SDGs. World Health Organization. 2018. URL: https://www.who.int/data/gho/publications/world-health-statistics [accessed 2019-03-20]
3. Statistics of China's health development in 2019. National Health Commission, PR China. 2020 Jun 8. URL: https://www.med66.com/yishizigekaoshi/jiaodian/mu2006088640.shtml [accessed 2020-09-28]
4. Hong Y. China's basic medical insurance system nearly covers entire population. People's Daily Online. URL: http://en.people.cn/n3/2020/0119/c90000-9650123.html [accessed 2020-10-01]
5. Critical illness insurance broke out and the number of insurance policies increased to 300 million! The products are dazzling, and the market still has two major shortcomings. SoHu. 2019 Oct 28. URL: https://www.sohu.com/a/350034425_177992 [accessed 2020-09-20]
6. What is the medical insurance reimbursement ratio in 2019, and how is it reimbursed? Policies are here. SoHu. 2019 May 17. URL: https://www.sohu.com/a/314528051_492006 [accessed 2020-07-01]

7. Seven million Chinese became poor due to illness, one was seriously ill and the whole family was sentenced. SoHu. 2018 Aug 27. URL: https://www.sohu.com/a/250259909_679841 [accessed 2019-09-03]
8. McClanahan C. People are raising $650 million on GoFundMe each year to attack rising healthcare costs. Forbes. 2018 Aug 13. URL: https://www.forbes.com/sites/carolynmcclanahan/2018/08/13/using-gofundme-to-attack-health-care-costs/#2d1215dd2859 [accessed 2019-12-01]
9. About Us (Introduction to Shuidichou). Shuidichou. 2018. URL: https://www.shuidichou.com/about [accessed 2018-10-10]
10. Burtch G, Chan J. Investigating the relationship between medical crowdfunding and personal bankruptcy in the United States: evidence of a digital divide. MISQ 2019 Jan 1;43(1):237-262 [FREE Full text] [doi: 10.25300/misq/2019/14569]
11. Gao F, Li X, Cheng Y, Hu Y. Ladies first, gentlemen third! The effect of narrative perspective on medical crowdfunding. SSRN J 2019 Apr 22:MKG-2019-1336. [doi: 10.2139/ssrn.3376215]
12. Jin P. Medical crowdfunding in China: empirics and ethics. J Med Ethics 2019 Aug;45(8):538-544. [doi: 10.1136/medethics-2018-105054] [Medline: 31217228]
13. 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]
14. Howard J. Crowdfunding raises millions for risky medical treatments, study says. CNN. 2018 Oct 23. URL: https://edition.cnn.com/2018/10/23/health/medical-fundraising-unproven-treatments-study/index.html [accessed 2019-10-20]
15. Guo X, Jiang J, Yang Y. 150,000 data interpretation of serious illness crowdfunding: Who can raise life? Sina. 2019 Jan 31. URL: http://finance.sina.com.cn/roll/2019-01-31/doc-ihqfskcp1985082.shtml?qq-pf-to=pcqq.c2c [accessed 2020-05-21]
16. Gonzales A, Kwon E, Lynch T, Fritz N. “Better everyone should know our business than we lose our house”: Costs and benefits of medical crowdfunding for support, privacy, and identity. New Media Soc 2016 Sep 07;20(2):641-658. [doi: 10.1177/1461444816667723]
17. Moore B. Medical crowdfunding and the virtuous donor. Bioethics 2019 Feb;33(2):238-244. [doi: 10.1111/bioe.12527] [Medline: 30341924]
18. Snyder J, Crooks VA, Mathers A, Chow-White P. Appealing to the crowd: ethical justifications in Canadian medical crowdfunding campaigns. J Med Ethics 2017 Jun;43(6):364-367 [FREE Full text] [doi: 10.1136/medethics-2016-103933] [Medline: 28137998]
19. Zenone M, Snyder J. Fraud in medical crowdfunding: a typology of publicized cases and policy recommendations. Policy Internet 2018 Oct 28;11(2):215-234. [doi: 10.1002/poi3.188]
20. Majumdar A, Bose I. My words for your pizza: An analysis of persuasive narratives in online crowdfunding. Inf Manag 2018 Sep;55(6):781-794. [doi: 10.1016/j.im.2018.03.007]
21. Tirdatov I. Web-based crowd funding: rhetoric of success. Tech Commun 2014 Feb;61(1):3-24.
22. Wang X, Wu Y, Guo J, Yang L. Says what in your photos? The impacts of photographic narratives on medical crowdfunding performance. 2020 Jun 4 Presented at: PACIS 2020; June 20-24, 2020; Dubai, UAE p. 231 URL: https://aisel.aisnet.org/pacis2020/231/
23. Durand WM, Peters JL, Eltorai AEM, Kalagara S, Osband AJ, Daniels AH. Medical crowdfunding for organ transplantation. Clin Transplant 2018 Jun;32(6):e13267. [doi: 10.1111/ctr.13267] [Medline: 29683220]
24. Xu K, Wang X. "Kindhearted People, Please Save My Family": Narrative strategies for new media medical crowdfunding. Health Commun 2020 Nov;35(13):1605-1613. [doi: 10.1080/10410236.2019.1654173] [Medline: 31455114]
25. Belleflamme P, Lambert T, Schwienbacher A. Crowdfunding: Tapping the right crowd. J Bus Ventur 2014 Sep;29(5):585-609. [doi: 10.1016/j.jbusvent.2013.07.003]
26. Li Y, Wu J, Hsieh C, Liou J. A social fundraising mechanism for charity crowdfunding. Decis Support Syst 2020 Feb;129:113170. [doi: 10.1016/j.dss.2019.113170]
27. Du Z, Li M, Wang K. “The more options, the better?” Investigating the impact of the number of options on backers’ decisions in reward-based crowdfunding projects. Inf Manag 2019 Apr;56(3):429-444. [doi: 10.1016/j.im.2018.08.003]
28. Lukkarinen A, Teich J, Wallenius H, Wallenius J. Success drivers of online equity crowdfunding campaigns. Decis Support Syst 2016 Jul;87:26-38. [doi: 10.1016/j.dss.2016.04.006]
29. Berns J, Figueroa-Armijos M, da Motta Veiga SP, Dunne T. Dynamics of lending-based prosocial crowdfunding: using a social responsibility lens. J Bus Ethics 2018 Jun 4;161(1):169-185. [doi: 10.1007/s10551-018-3932-0]
30. Gleasure R, Feller J. Does heart or head rule donor behaviors in charitable crowdfunding markets? Int J Electron Commer 2016 Jun 20;20(4):499-524. [doi: 10.1080/10864415.2016.1171975]
31. Renwick MJ, Mossialos E. Crowdfunding our health: Economic risks and benefits. Soc Sci Med 2017 Oct;191:48-56 [FREE Full text] [doi: 10.1016/j.socscimed.2017.08.035] [Medline: 28889030]
32. 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]
33. Murdoch B, Marcon AR, Downie D, Caulfield T. Media portrayal of illness-related medical crowdfunding: A content analysis of newspaper articles in the United States and Canada. PLoS One 2019;14(4):e0215805 [FREE Full text] [doi: 10.1371/journal.pone.0215805] [Medline: 31013310]

34. Greenberg M, Pardo B, Hariharan K, Gerber E. New York, NY: Association for Computing Machinery; 2013 Apr Presented at: CHI'13 Extended Abstracts on Human Factors in Computing Systems; April 2013; Paris, France p. 1815-1820. [doi: 10.1145/2468356.2468682]
35. Lin M. Economic value of texts: evidence from online debt crowdfunding. Federalreserve.gov. 2016 Dec 2. URL: https://www.federalreserve.gov/conferences/files/economic-value-of-texts-presentation.pdf [accessed 2019-11-09]
36. Kshetri N. Success of crowd-based online technology in fundraising: an institutional perspective. J Int Manag 2015 Jun;21(2):100-116. [doi: 10.1016/j.intman.2015.03.004]
37. Xu L. Will a digital camera cure your sick puppy? Modality and category effects in donation-based crowdfunding. Telemat Informat 2018 Oct;35(7):1914-1924. [doi: 10.1016/j.tele.2018.06.004]
38. Zubrickas R. The provision point mechanism with refund bonuses. J Public Econ 2014 Dec;120:231-234. [doi: 10.1016/j.jpubeco.2014.10.006]
39. Cason T, Zubrickas R. Donation-based crowdfunding with refund bonuses. Eur Econ Rev 2019 Oct;119:452-471. [doi: 10.1016/j.euroecorev.2019.08.003]
40. Sasaki S. Majority size and conformity behavior in charitable giving: Field evidence from a donation-based crowdfunding platform in Japan. J Econ Psychol 2019 Jan;70:36-51. [doi: 10.1016/j.joep.2018.10.011]
41. Cheng F, Chen T, Chen Q. Matching donations based on social capital in Internet crowdfunding can promote cooperation. Physica A 2019 Oct;531:121766. [doi: 10.1016/j.physa.2019.121766]
42. Aristotle, Roberts W. The rhetoric and the poetics of Aristotle. New York City, NY: Random House; 1954.
43. Hong W, Yu Z, Wu L, Pu X. Influencing factors of the persuasiveness of online reviews considering persuasion methods. Electron Commer Res Appl 2020 Jan;39:100912. [doi: 10.1016/j.elerap.2019.100912]
44. Aleksina A, Akulenka S, Lublóy Á. Success factors of crowdfunding campaigns in medical research: perceptions and reality. Drug Discov Today 2019 Jul;24(7):1413-1420. [doi: 10.1016/j.drudis.2019.05.012] [Medline: 31132416]
45. Kusumarani R, Zo H. Why people participate in online political crowdfunding: A civic voluntarism perspective. Telemat Informat 2019 Aug;41:168-181. [doi: 10.1016/j.tele.2019.04.011]
46. Liang T, Wu S, Huang C. Why funders invest in crowdfunding projects: Role of trust from the dual-process perspective. Inf Manag 2019 Jan;56(1):70-84. [doi: 10.1016/j.im.2018.07.002]
47. Petitjean M. What explains the success of reward-based crowdfunding campaigns as they unfold? Evidence from the French crowdfunding platform KissKissBankBank. Finance Res Lett 2018 Sep;26:9-14. [doi: 10.1016/j.frl.2017.11.005]
48. Kim T, Por M, Yang S. Winning the crowd in online fundraising platforms: The roles of founder and project features. Electron Commer Res Appl 2017 Sep;25:86-94. [doi: 10.1016/j.elerap.2017.09.002]
49. Allison T, Davis B, Webb J, Short J. Persuasion in crowdfunding: An elaboration likelihood model of crowdfunding performance. J Bus Ventur 2017 Nov;32(6):707-725. [doi: 10.1016/j.jbusvent.2017.09.002]
50. Bi S, Liu Z, Usman K. The influence of online information on investing decisions of reward-based crowdfunding. J Bus Res 2017 Feb;71:10-18. [doi: 10.1016/j.jbusres.2016.10.001]
51. Gorbatai A, Nelson L. Gender and the language of crowdfunding. Acad Manag Proc 2015 Jan;2015(1):15785 [FREE Full text] [doi: 10.5465/ambpp.2015.15785abstract]
52. Dong Z, Dong Q, Hao C. The Nature of HowNet. In: HowNet and the Computation of Meaning. Singapore: World Scientific; Feb 2006:17-19.
53. Fu X, Liu W, Xu Y, Cui L. Combine HowNet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis. Neurocomputing 2017 Jun;241:18-27. [doi: 10.1016/j.neucom.2017.01.079]
54. Zhang W, Xu H, Wan W. Weakness finder: find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Syst Appl 2012 Sep;39(11):10283-10291. [doi: 10.1016/j.eswa.2012.02.166]
55. FunInFunding. URL: https://qsgy.qschou.com/ [accessed 2020-06-01]
56. Iyengar S, McGuire WJ. Explorations in Political Psychology. Durham, NC: Duke University Press; Jun 1993:357-358.
57. Ren J, Raghupathi V, Raghupathi W. Understanding the dimensions of medical crowdfunding: a visual analytics approach. J Med Internet Res 2020 Jul 03;22(7):e18813 [FREE Full text] [doi: 10.2196/18813] [Medline: 32618573]
58. Sisco MR, Weber EU. Examining charitable giving in real-world online donations. Nat Commun 2019 Sep 03;10(1):3968. [doi: 10.1038/s41467-019-11852-z] [Medline: 31481650]
59. Wymer W, Drollinger T. Charity appeals using celebrity endorsers: celebrity attributes most predictive of audience donation intentions. Voluntas 2014 Dec 10;26(6):2694-2717. [doi: 10.1007/s11266-014-9546-y]
60. Behl A, Dutta P. Engaging donors on crowdfunding platform in Disaster Relief Operations (DRO) using gamification: A Civic Voluntary Model (CVM) approach. Int J Inf Manag 2020 Oct;54:102140. [doi: 10.1016/j.ijinfomgt.2020.102140]

Edited by R Kukafka; submitted 10.07.20; peer-reviewed by J Mou, T Mattson; comments to author 11.09.20; revised version received 05.11.20; accepted 25.04.21; published 11.06.21.

Please cite as:
Zhang F, Xue B, Li Y, Li H, Liu Q
Effect of Textual Features on the Success of Medical Crowdfunding: Model Development and Econometric Analysis from the Tencent Charity Platform
J Med Internet Res 2021;23(6):e22395
URL: https://www.jmir.org/2021/6/e22395
doi:10.2196/22395
PMID:34114959

Original Paper

Exploratory Outlier Detection for Acceleromyographic Neuromuscular Monitoring: Machine Learning Approach

Michaël Verdonck[1*], PhD; Hugo Carvalho[2*], MD, MSc; Johan Berghmans[3], MD, PhD; Patrice Forget[4], MD, PhD; Jan Poelaert[2], MD, PhD

[1]Department of Anesthesiology and Perioperative Medicine, Vrije Universiteit Brussel, Jette, Belgium
[2]Department of Anesthesiology and Perioperative Medicine, Universitair Ziekenhuis Brussel, Brussel, Belgium
[3]Department of Anesthesiology, Ziekenhuis Netwerk Antwerpen, Antwerp, Belgium
[4]Department of Anaesthesia, University of Aberdeen, Aberdeen, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Michaël Verdonck, PhD
Department of Anesthesiology and Perioperative Medicine
Vrije Universiteit Brussel
Laarbeeklaan 101
Jette, 1050
Belgium
Phone: 32 474683824
Email: michael.verdonck@vub.be

Abstract

Background: Perioperative quantitative monitoring of neuromuscular function in patients receiving neuromuscular blockers has become internationally recognized as an absolute and core necessity in modern anesthesia care. Because of their kinetic nature, artifactual recordings of acceleromyography-based neuromuscular monitoring devices are not unusual. These generate a great deal of cynicism among anesthesiologists, constituting an obstacle toward their widespread adoption. Through outlier analysis techniques, monitoring devices can learn to detect and flag signal abnormalities. Outlier analysis (or anomaly detection) refers to the problem of finding patterns in data that do not conform to expected behavior.

Objective: This study was motivated by the development of a smartphone app intended for neuromuscular monitoring based on combined accelerometric and angular hand movement data. During the paired comparison stage of this app against existing acceleromyography monitoring devices, it was noted that the results from both devices did not always concur. This study aims to engineer a set of features that enable the detection of outliers in the form of erroneous train-of-four (TOF) measurements from an acceleromyographic-based device. These features are tested for their potential in the detection of erroneous TOF measurements by developing an outlier detection algorithm.

Methods: A data set encompassing 533 high-sensitivity TOF measurements from 35 patients was created based on a multicentric open label trial of a purpose-built accelero- and gyroscopic-based neuromuscular monitoring app. A basic set of features was extracted based on raw data while a second set of features was purpose engineered based on TOF pattern characteristics. Two cost-sensitive logistic regression (CSLR) models were deployed to evaluate the performance of these features. The final output of the developed models was a binary classification, indicating if a TOF measurement was an outlier or not.

Results: A total of 7 basic features were extracted based on raw data, while another 8 features were engineered based on TOF pattern characteristics. The model training and testing were based on separate data sets: one with 319 measurements (18 outliers) and a second with 214 measurements (12 outliers). The F1 score (95% CI) was 0.86 (0.48-0.97) for the CSLR model with engineered features, significantly larger than the CSLR model with the basic features (0.29 [0.17-0.53]; *P*<.001).

Conclusions: The set of engineered features and their corresponding incorporation in an outlier detection algorithm have the potential to increase overall neuromuscular monitoring data consistency. Integrating outlier flagging algorithms within neuromuscular monitors could potentially reduce overall acceleromyography-based reliability issues.

Trial Registration: ClinicalTrials.gov NCT03605225; https://clinicaltrials.gov/ct2/show/NCT03605225

(J Med Internet Res 2021;23(6):e25913) doi:10.2196/25913

KEYWORDS

neuromuscular monitoring; outlier analysis; acceleromyography; postoperative residual curarization; train-of-four; monitoring devices; neuromuscular; machine learning; monitors; anesthesiology

Introduction

Postoperative residual curarization remains a frequent and often concealed event within modern anesthesia care [1]. It translates clinically into complications such as aspiration of gastric contents [2,3] and an impaired ventilatory response to hypoxia [4]. This is ultimately linked to an increase of morbidity and mortality due to postoperative pulmonary complications [5]. As such, perioperative quantitative monitoring of neuromuscular function in patients receiving neuromuscular blockers has become internationally recognized as an absolute and core necessity in modern anesthesia care [6,7]. Besides a reduction of the incidence of severe respiratory complications [8-10], quantitative monitoring also potentially leads to considerable financial health care savings, with complications stemming from suboptimal neuromuscular monitoring being estimated to be as high as US $25.000 per patient per event [7].

Although a seemingly straightforward procedure, neuromuscular monitoring presents users with nuances that are frequently overlooked or that are prone to misinterpretation [11]. This has been exemplified by research [12,13] showing that baseline (control) train-of-four ratios (TOFRs; T4/T1) at the adductor pollicis frequently assume supra-physiological values (TOFR > 1) when measured using acceleromyography (AMG). Similarly, Kopman et al [13] have scrutinized some algorithmic simplifications used by common AMG monitors (T4/T2 ratio as a substitute for T4/T1) and how their validity is dependent on the degree of recovery from nondepolarizing neuromuscular block. Such interpretative considerations, associated frequent artifactual confounders, and known overestimation tendencies when compared with electromyography (EMG) or mechanomyography [14] contribute to the perpetuation of anesthesiologist's cynicism toward objective neuromuscular monitoring methods, further hindering their widespread adoption [15].

The herein presented research has been motivated by the development of a smartphone app intended for neuromuscular monitoring based on combined accelerometric and angular hand movement data [16]. During the paired comparison stage of this app against existing AMG monitoring devices, it was noted that the results from both devices did not always concur. For instance, it was observed that the collected raw movement data regularly displayed nonstandard TOF patterns, whereas the AMG neuromuscular monitoring device did not appear to detect these outliers and displayed a seemingly (oversimplified) TOFR plotting. As with any instrument that aims to measure a certain signal [17], the measurement of TOFR is similarly prone to the appearance of outliers, which can be erroneously interpreted as correct measurements.

From a data analysis standpoint, outlier analysis techniques can be adopted to increase data reliability. Outlier analysis (or anomaly detection) refers to the problem of finding patterns in data that do not conform to expected behavior [18]. This study aims to conduct an offline exploratory analysis on raw AMG neuromuscular monitoring data and to engineer features (or variables) to be able to flag erroneous TOF measurements. These features will be subsequently evaluated for their usability in outlier analysis.

Methods

Overview

This manuscript follows the "Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View" [19]. The key steps of the feature engineering and the development of the outlier analysis algorithm are summarized below.

Data Recruitment and Preprocessing

All data for this study have been collected during a prospective open-label bicentric clinical trial (Clinical Trial Identifier NCT03605225) that took place in Ziekenhuisnetwerk Antwerpen Middelheim (Antwerp, Belgium) and Universitair Ziekenhuis Brussel (Brussels, Belgium). Registration occurred prior to the start of the trial. Data collection started in February 2018 and terminated in April 2019. The trial was conducted in accordance with the established protocol after approval by the Medical Ethical Committees of both hospitals (ZNA Middelheim reference number 5055; UZBrussel reference 2018/031, BUN 009201835039). It followed current good clinical practice guidelines and applicable law(s), as well as adhered to the applicable CONSORT guidelines.

The data used for the algorithm development were collected using a purpose-built smartphone app specifically aimed to monitor hand movements evoked by extraneural supramaximal stimulation of the ipsilateral nervus ulnaris by means of a peripheral nerve stimulator. Collected data included triaxial (3D) raw acceleration values (m/s^2) as well as raw 3D angular velocity values (rad s^{-1}).

Earlier trials involving beta versions of this app have been published and they reported bidirectional 95% limits of agreement of 0.12 (TOFR, absolute units) when compared with a standard AMG neuromuscular CE/FDA-labeled monitor [16]. This study included 35 patients, with a total of 533 TOF measurements. The offline evaluation of the observations was performed by 2 authors of this paper (MV and HC), who performed the evaluation independently from one another. The classification of outliers was afterward compared and corresponded to a 98% agreement on the labeling of TOF patterns. In total, 30 of the 533 observations were identified as outliers. These anomalies were detected in the TOF measurements of 18 patients.

Feature Engineering

The acceleration and angulation signals were collected through the open-source Cordova Plugin Device-motion library [20], and measure the movement of the muscle contractions in the 3 orthogonal directions of movement (X, Y, and Z). Similar to studies in mechanomyography [21], root-mean square analysis was performed on these signals to indicate the range of muscle displacement represented by its acceleration (expressed with units of m/s^2; Figure 1). The features for the model were then directly derived from the continuous monitoring signal. A first basic set of features (Table 1) were derived: the combined acceleration value of multiple points on the vicinity of the peak of the different TOF twitches (T1, T2, T3, T4) and the absolute TOFR value.

Figure 1. Normal and anomalous train-of-four (TOF) patterns; upper figures display normal TOF observations while bottom figures represent anomalies. Normality recordings are illustrated by the upper 2 patterns, where 4 clear peaks can be detected and that follow one another in a fixed time interval. The lower 2 patterns represent 2 simulated anomalies, where the 4 peaks cannot be clearly recognized from the TOF pattern; or where a wide gap in terms of time interval exists between peaks.

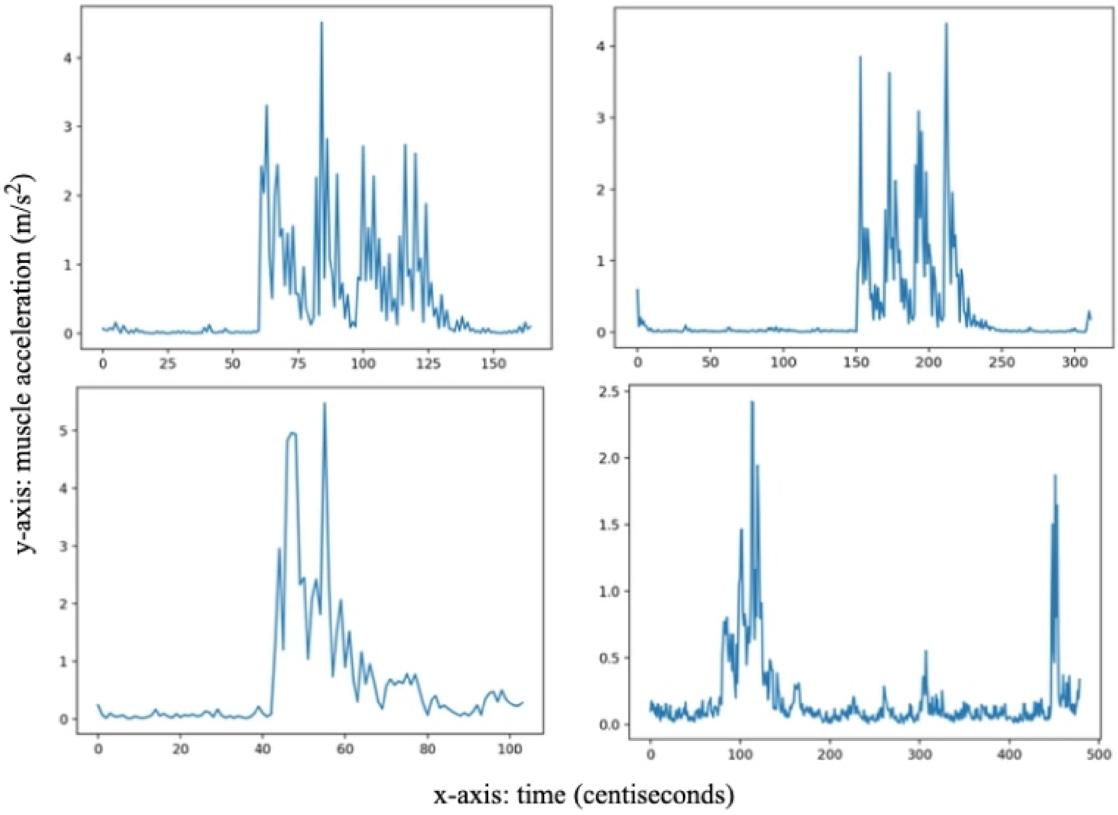


Table 1. Description of the basic features of data set. A train-of-four recording is defined as the integral combination of all acceleration/angulation points of T1, T2, T3, and T4.

Feature name	Description	n (count)
T1	First twitch of train-of-four response	533
T2	Second twitch of train-of-four response	533
T3	Third twitch of train-of-four response	533
T4	Fourth twitch of train-of-four response	533
TOFR[a]	Absolute ratio derived by dividing T1 with T4	533
AMG[b]_StdDev	Standard deviation of an AMG measurement	533
AMG_Mean	Arithmetic mean of an AMG measurement	533

[a] TOFR: train-of-four ratio.

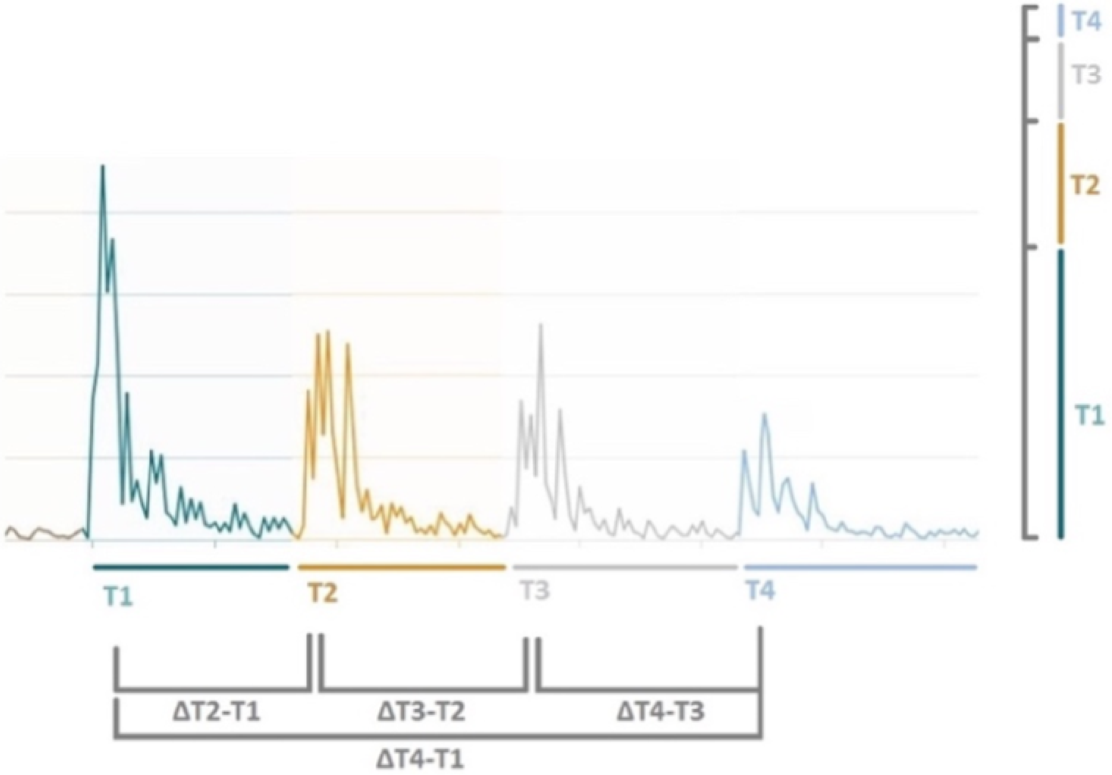
[b] AMG: acceleromyography.

Additionally, the arithmetic mean and standard deviation of the AMG values related to one measurement were computed in order to gain a better insight into the differences in variation between different AMG measurements. In order to avoid confusion with other descriptive statistics, these were labeled as “AMG_StdDev” and “AMG_Mean.” An additional set of features were engineered to assess specific TOF pattern characteristics (Table 2). Several of these were derived from the distance between the different TOF twitches (denominated with the prefix “delta”; Figure 2), whereas other features were based on the ratio of a specific TOF twitch compared with the mean of the respective TOF measurement (denominated with the prefix “ratio”). With a nonanomalous pattern, one would logically expect such a ratio to be exceedingly higher than the arithmetic mean of the TOF measurement. In total, 15 features were extracted and engineered to serve as input for the outlier detection models. We emphasize that the authors have deliberately designed the aforementioned features to detect an anomaly based on the signal characteristics of a specific measurement instance. Because of the limited size of the data set, there is no feature that takes into account any historical information of the TOF recording, administered drugs, or other patient-related parameters.

Table 2. Description of engineered features of data set.

Feature name	Description	n (count)
deltaT2_T1	Elapsed time (ms) between second and first twitch	533
deltaT3_T2	Elapsed time (ms) between third and second twitch	533
deltaT4_T3	Elapsed time (ms) between fourth and third twitch	533
deltaT4_T1	Elapsed time (ms) between fourth and first twitch	533
ratioT1	Ratio between first twitch and arithmetic mean	533
ratioT2	Ratio between second twitch and arithmetic mean	533
ratioT3	Ratio between third twitch and arithmetic mean	533
ratioT4	Ratio between fourth twitch and arithmetic mean	533

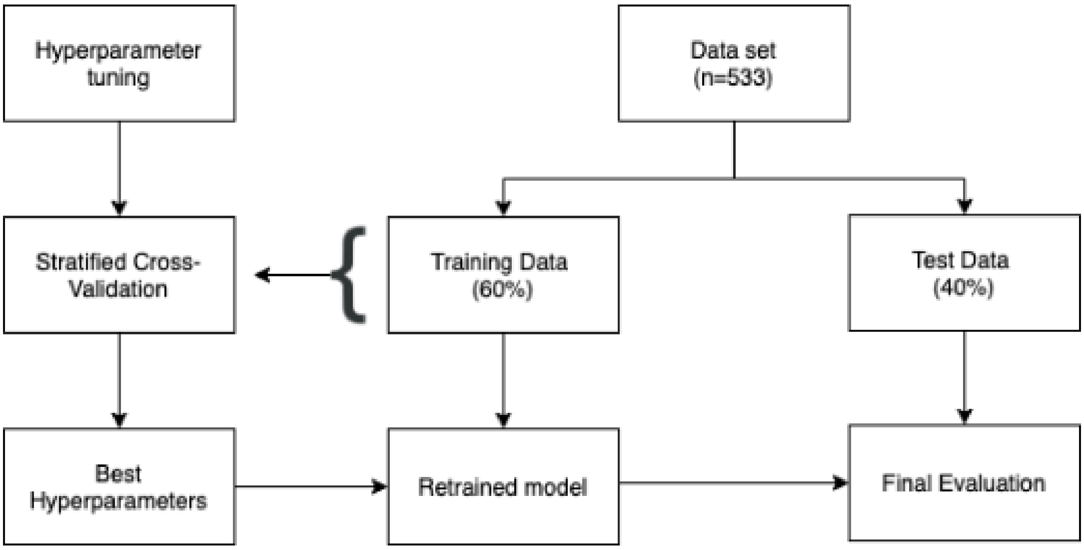
Figure 2. Basic and engineered feature illustration (x axis: time, y axis: combined angulation and acceleration). Each color represents an individual train-of-four (TOF) twitch. Each individual twitch (T1, T2, T3, and T4) is composed of multiple acceleration/angulation points during the corresponding contraction, and not solely by the highest value. The TOF recording is obtained by the summation of each individual twitch.



Model Development and Overfitting

Model development encompassed both the basic feature set and the engineered feature set. Because the study’s data set is composed out of labeled data with 2 distinct classes (normal observations and outliers), supervised learning can be applied in the form of a classification model. Because outliers are rare instances in the data, there is a class imbalance where the distribution between the normal observations and outliers is significantly skewed. To overcome the issue of class imbalance, a cost-sensitive learning technique was adopted, where the objective function of the classification algorithm is modified in order to weight the classification errors in a differential way for the normal and the less frequent class. This refers concretely to cost-sensitive logistic regression (CSLR), where a class weighting configuration is used to influence the amount of logistic regression coefficients that are updated during training. The weighting penalizes the model less for errors made on instances from the normal class, while maintaining a larger penalty for errors made on instances from the rare class. The result is a version of logistic regression that performs better on imbalanced classification tasks [17]. To avoid overfitting, a part of the available data set was held out as a test set during the data preprocessing phase (Figure 3). Moreover, in order to select the different hyperparameters related to our chosen models, a cross-validation strategy was adopted, more specifically in the form of stratified k-fold validation together with the hyperparameter optimization technique called grid search.

Figure 3. Overview of data set partitioning and model training.



Model Evaluation

To evaluate the effectiveness of the engineered features (Table 2), their performance was compared with that of the basic features (Table 1) and tested for significance. Therefore, 2 CSLRs were trained: one on the basic feature set and another on the engineered ones. All model performances were assessed on the same test set, which was composed of 40% of the total data and separated from the training set. The precision, recall, and the F1 score were chosen as performance evaluation metrics as these are best suited to evaluate data sets with class imbalance. Additionally, receiver operating characteristics (ROCs) and area under the curve (AUC) graphs were computed to characterize the performance of the models.

Statistical Analysis

Statistical analysis was performed with the open-source python library Scikit-learn [22]. Wilcoxon signed-rank test was adopted to compare the performance of the CSLR models, with $P<.05$ considered significant.

Results

Descriptive Statistics, Model Training, and Cross-Validation

The descriptive statistics of both feature sets are quantified in Table 3. The Python code related to hyperparameter optimization, training, and testing is presented in Multimedia Appendix 1. In contrast to the basic features, the engineered variables display a larger variation, as can be derived from their standard deviation. In fact, mainly the "delta" variables appear to be skewed. Figure 4 displays a scatter plot and a distribution plot of the basic features T1 and TOFR, and the engineered features ratioT1 and deltaT4_T1. While the outliers in the scatter plot of the basic features are more dispersed throughout the normal data, the outliers within the scatter plot of the engineered plots can more clearly be identified compared with the normal observations.

Concerning model training, the training data set (n=319) consisted out of 18 outliers, while the test data set (n=214) included 12 outliers. Both train and test data instances were chosen in a completely randomized manner by means of the train_test_split function of the scikit-learn library [22]. The division of data during cross-validation was performed solely on the segmented measurements (n=533). The split between train and test data has also been performed in a stratified way so as to guarantee the same class–imbalanced distribution of the entire data set. As for the stratified k-fold learning, 5 folds were chosen to split the training set, taking into consideration the size of the training set. In combination with cross validation, a grid search approach was employed to find the best hyperparameters for the L2 regularization coefficient and the appropriate class weights configuration for the imbalanced class distributions of our data set. All model training, cross-validation, and model evaluation were performed with the open-source library scikit-learn and the high-level programming language Python (version 3.8.2) [23].

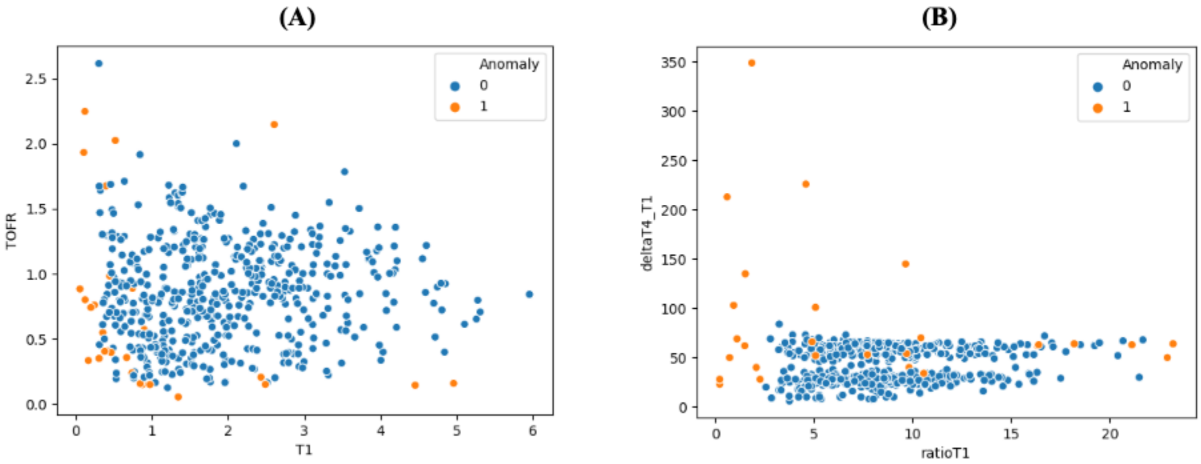
Table 3. Descriptive statistics of basic features and engineered features.

Features	Mean (SD)	25%	50%	75%	Minimum	Maximum	Kurtosis	Skewness
T1	1.89 (1.13)	0.06	0.97	1.67	2.62	5.96	0.17	0.77
T2	1.82 (1.2)	0.02	0.8	1.63	2.58	6.03	0.09	0.78
T3	1.65 (1.25)	0.04	0.58	1.36	2.38	6.78	0.77	1.04
T4	1.6 (1.25)	0.05	0.56	1.26	2.34	6.3	0.62	1.05
TOFR[a]	0.84 (0.4)	0.05	0.53	0.82	1.1	2.62	0.48	0.52
AMG[b]_StdDev	0.44 (0.32)	0.05	0.18	0.36	0.61	2.01	2.54	1.31
AMG_Mean	0.27 (0.23)	0.04	0.11	0.2	0.35	1.5	5.91	2.06
deltaT2_T1	13.99 (6.55)	1	9	12	20	60	4.78	0.97
deltaT3_T2	14.44 (7.23)	2	9	13	20	80	20.16	2.47
deltaT4_T3	16.3 (18.95)	2	9	14	20	332	164.56	11.27
deltaT4_T1	44.73 (25.36)	6	28	48	60	349	46.29	4.61
ratioT1	8.55 (3.73)	0.22	5.96	7.95	10.41	23.21	1.39	0.94
ratioT2	7.57 (2.83)	0.04	5.72	7.23	9.09	20.19	1.27	0.73
ratioT3	6.52 (2.76)	0.09	4.47	6.16	8.17	18.46	1.46	0.92
ratioT4	6.36 (2.9)	0.19	4.3	6.05	8.15	18.47	1.03	0.74

[a] TOFR: train-of-four ratio.

[b] AMG: acceleromyography.

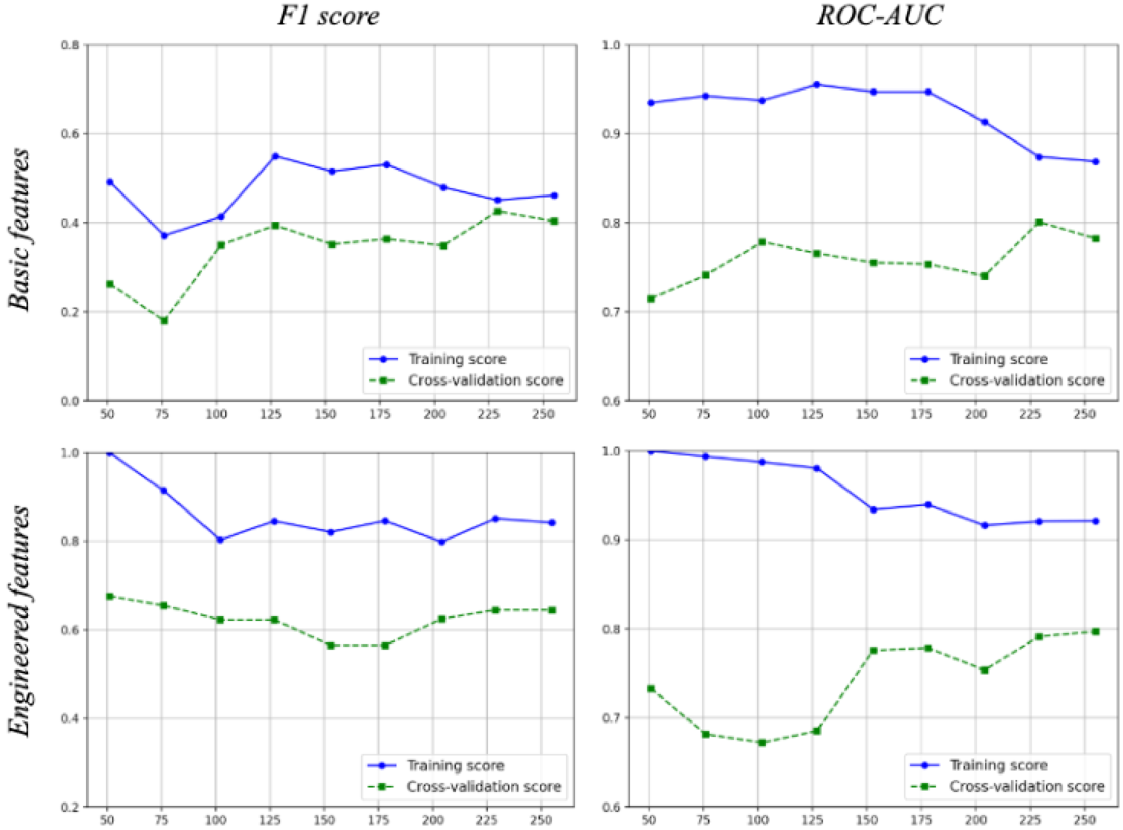
Figure 4. Panel A displays a scatter plot of train-of-four ratio (TOFR) and T1. Panel B displays a scatter plot of the features deltaT4_T1 and ratioT1. Scatter plot displays 1 as outlier, and 0 as a normal observation. TOFR in absolute units. DeltaT4_T1 in milliseconds. T1 in root-mean square angulation and acceleration.



Model Performance

Figure 5 presents the learning curves during model training and validation, for both the F1 score and ROC–AUC performance metrics. By plotting the model training and validation performances as functions of the training set size, high variance (ie, overfitting) or bias (ie, underfitting) can be assessed. While overall there seems to be a good bias–variance trade-off for both models, the CSLR based on the engineered features data set tends to overfit more than the model based on the basic features.

Figure 5. Learning curves of the cost-sensitive logistic regression models of the basic feature (above) set and the engineering feature set (below). Training and validation metrics are the F1-score and the ROC-AUC score (y-axis). X-axis represents the number of training instances.



In Table 4, the F1 score, the ROC–AUC, precision, and recall for the CSLRs of the basic features and the engineered features are presented. Performance metrics are given for the training and test data sets. The CSLR model with the engineered features on the test data has an improved performance compared with the metrics of the training data set, indicating that the model has not been overfit. For the CSLR model of the basic features, we observe the opposite. The F1 score (95% CI) was 0.86 (0.48-0.97) for the CSLR model with the engineered features, which was significantly larger than the CSLR model with the basic features (0.29 [0.17-0.53]; *P*<.001). ROC curves and AUC curve results are visualized in Figure 6. The CSLR model with the engineered features has the highest AUC (95% CI) with a score of 0.91 (0.72-0.97), significantly larger than the CSLR model with the basic features (0.86 [0.63-0.93]; *P*<.001).

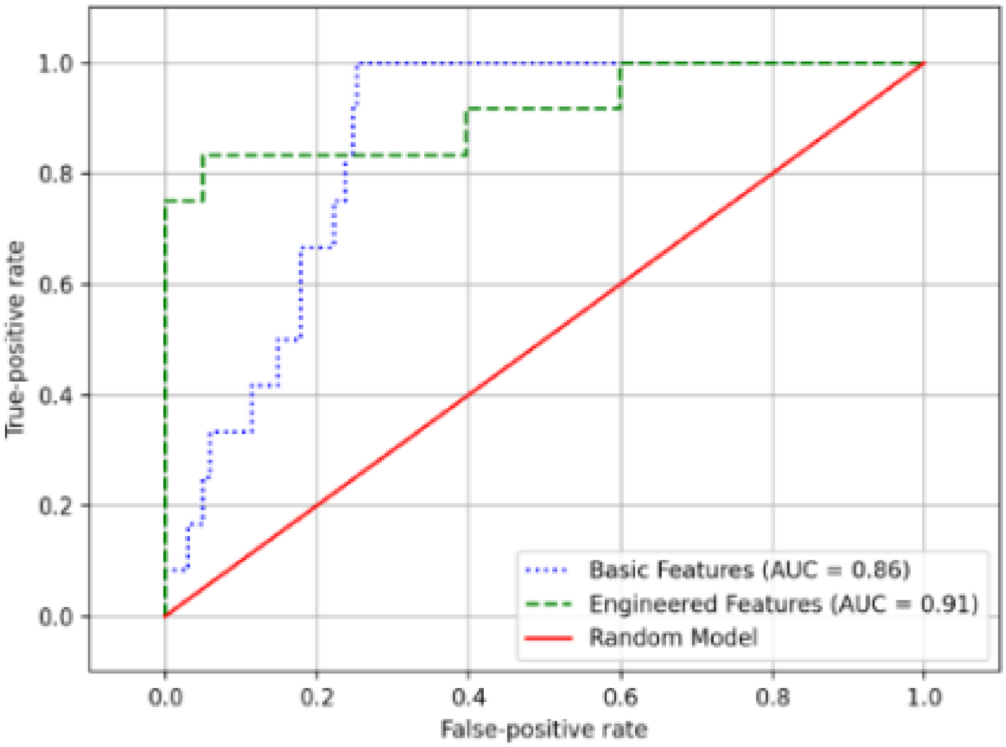
Table 4. Performance metrics of the training data set and the test data set.

Data sets	F1 score, mean (95% CI)	ROC[a]–AUC[b], mean (95% CI)	Precision, mean (95% CI)	Recall, mean (95% CI)
Training data set (n=319)				
Basic features	0.47 (0.24-0.63)	0.78 (0.63-0.82)	0.43 (0.18-0.68)	0.55 (0.33-0.71)
Engineered features	0.65 (0.49-0.84)	0.80 (0.70-0.87)	1.00 (0.68-1.00)	0.50 (0.40-0.75)
Test data set (n=214)				
Basic features	0.29 (0.17-0.53)	0.86 (0.63-0.93)	0.25 (0.10-0.52)	0.33 (0.29-0.98)
Engineered features	0.86 (0.48-0.97)	0.91 (0.72-0.97)	1.00 (0.49-1.00)	0.75 (0.44-0.94)

[a] ROC: receiver operating characteristic.

[b] AUC: area under the curve.

Figure 6. Receiver-operating characteristic curves displaying the ability of the algorithms to classify a train-of-four measurement as an outlier. AUC: area under the curve.



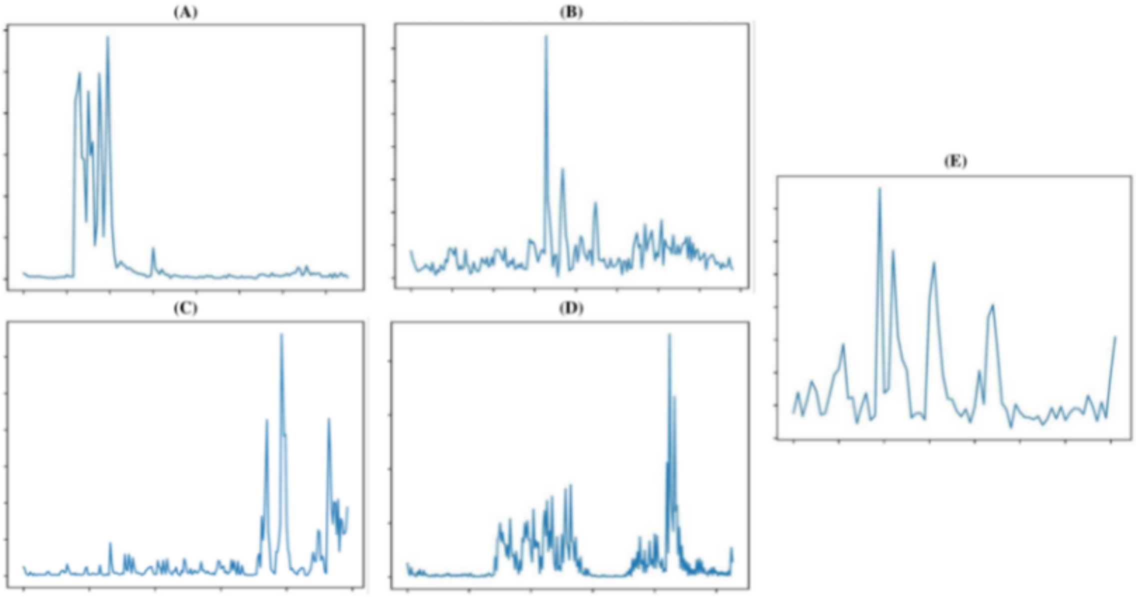
Discussion

Principal Findings

The herein obtained results demonstrate that engineered TOF features outperform basic and common clinically employed neuromuscular monitoring endpoints for automated outlier identification of intraoperative TOF measurements. In the test data set, the CSLR of the engineered variables correctly identified 9 out of 12 measurements as outliers, compared with the CSLR of the basic features, which only correctly flagged one-third of the outliers. Moreover, the basic feature CSLR displayed a high degree of false positives, where 12 TOF measurements were incorrectly labeled as outliers, as opposed to 0 false positives on the CSLR model with the engineered features. While the authors recognize that the current models are sedimented on a limited data set and that further development is necessary in order to arrive at a clinically deployable outlier detection algorithm, the performance of the engineered feature algorithm is promising for a possible clinical application. Various research efforts [8,15,24] have highlighted that quantitative neuromuscular monitoring is still suboptimally and reluctantly adopted by practicing anesthesiologists, among others, due to a low perceived usefulness and reliability of monitoring devices. Even when effectively available on request, perceived unreliability due to artifactual recordings has been shown to be a prevalent barrier and a technical hindrance toward consistent monitoring adoption [15]. Nevertheless, the measurement error (artifact) incidence rates of standard neuromuscular monitors are unknown, and for that matter, so is an encompassing formal description and corresponding physiological correlations.

Given the wide scope of anesthesia monitoring, the daily clinical relevance and eventual successful adoption of such an outlier analysis are certainly subject to debate. Nevertheless, the development of the CSLR models within this study is anchored on clinically grounded reported monitoring issues [7,8,15,25-29]. Figure 7 illustrates examples of flagged abnormal extraneural stimulation-induced moments. Although all have quantifiable and within-normality TOFR values, outliers are evident on closer inspection.

Figure 7. (A) Short 4-peak burst—rebound phenomenon TOF-like pattern after a single movement, short intertwitch distance; (B) T3 less than T2 and T4; (C) Almost equidistant oscillations; (D) Crescendo pattern; (E) Decrescendo but gross oscillations and variable interpeak distance. Some of the detected patterns have implications. y axis: muscle acceleration (m/s2), x axis: time (centiseconds). TOF: train-of-four.



The anchoring of such outlier analysis to the clinical context of neuromuscular blockade monitoring is yet to be done practically and prospectively. Although the authors anticipate the present offline analysis will improve both neuromuscular monitoring adoption and clinical errors when embedded into anesthesia monitors, this can only be speculated upon at the present stage of development. It should additionally be reiterated that the bottleneck issue of undereducation is not tackled by the present developments. In fact, phenomena such as the failure of T1% to reach its baseline levels of around 100% during EMG-based neuromuscular monitoring are frequently observed in clinical practice and similarly trigger distrust among anesthesiologists [15]. The importance of human factors on the effective implementation of recommendation software has been highlighted by a recent randomized pilot trial, where a rather widely deployable predictive algorithm—the Hypotension Prediction Index (Edwards Lifesciences Corporation)—has been shown to fail to engage anesthesiologists [30].

As reinforced in the latest perioperative neuromuscular management consensus statement, educational efforts constitute an important part of modern anesthetic neuromuscular monitoring [7]. On the authors' opinion, automated decision support software alone is expected to aid, but not solve or abolish, with the problem of suboptimal worldwide adoption of neuromuscular monitoring.

Considering the frequent and known artifactual biasing of kinetic data even with CE/FDA-labeled AMG/kinemyography/EMG neuromuscular monitoring devices [25-29], the added value of outlier analysis becomes especially relevant for reliability purposes. For instance, Liang et al [29] performed an ipsilateral comparison of AMG and EMG monitoring devices, concluding that AMG is less precise than EMG and overestimates the EMG TOFR by at least 0.15 units. A similar study performed by Kopman et al [28] found that AMG TOF values tend to overestimate the extent of EMG recovery, with a bias estimate of 0.125. However, both authors could not provide a cause to why these bias estimates between AMG and EMG devices were measured, stating that their lack of agreement cannot be explained by the imprecision of either device. Although outliers are not expected to explain such systematically reported intermethod precision differences when no technical issues are at hand, these do have the potential to compensate for small kinetic nuances such as overshoots, provided enough training data are available. In that sense, commonly applied AMG correction techniques such as normalization against baseline measurements could potentially be obviated. Nevertheless, based on the herein presented results, such potential is purely speculative.

Study Limitations

- It can be argued that the data set that was collected and applied to develop the algorithm within this study is of a somewhat limited sample size. In order to address this limitation and avoid overfitting, we chose a cost-sensitive learning technique for logistic regression, adopted a cross-validation strategy together with the grid search optimization technique, and implemented regularization training, commonly used in machine learning.
- While the presented CSLR model with our engineered features is capable of detecting outliers in the process of AMG neuromuscular monitoring, there is no correlation with a possible cause. Hence, except for a measurement repetition, it remains unclear how clinicians would act upon the warning given by the algorithm.
- The data of this study were collected through an AMG-based smartphone app specifically aimed to monitor the TOF movement pattern and to calculate the corresponding TOFR. This app is undergoing further refinement to facilitate neuromuscular monitoring and to provide clinical intraoperative decision support. While the quality of the recorded measurements of the device has been previously assessed within a clinical trial [16], it is

still possible that certain outliers are due to the nature of handheld devices themselves.

- We present a set of engineered features that have the potential for real-time detection of outliers within neuromuscular monitoring. Further analysis could reveal features that carry additional information for this purpose. Additionally, this study did not analyze (online) real-time streams of data to detect outliers.
- The algorithm developed depends on AMG neuromuscular monitoring devices. While these devices are still the most adopted quantitative neuromuscular devices in the domain of anesthesia [6], the developed algorithm does not tackle EMG-based devices [31].
- This study is not an intermethod validation study, but a precision increasing exercise that still warrants prospective intermethod comparison. This refers to its paired comparison with both AMG and EMG devices.

Conclusion

This study demonstrates that a set of engineered features has the ability to detect outliers from an AMG neuromuscular device based on intraoperative measurements. The development of the model based on these features displayed promising results toward the creation of an outlier detection technique for neuromuscular monitoring. With further research and additional training, an outlier detection algorithm can potentially be implemented within an AMG neuromuscular monitoring device to scan TOF measurements for outliers automatically while not relying on active input from medical providers.

Acknowledgments

This research was funded by the Flanders Innovation and Entrepreneurship Fund, the Willy Gepts Fund for Scientific Research, and the Society for Anesthesia and Resuscitation of Belgium (SARB).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Jupyter Notebook including Python code of cost-sensitive logistic regression classifier.

[ZIP File (Zip Archive), 2 KB - jmir_v23i6e25913_app1.zip]

References

1. Carvalho H, Verdonck M, Cools W, Geerts L, Forget P, Poelaert J. Forty years of neuromuscular monitoring and postoperative residual curarisation: a meta-analysis and evaluation of confidence in network meta-analysis. Br J Anaesth 2020 Oct;125(4):466-482 [FREE Full text] [doi: 10.1016/j.bja.2020.05.063] [Medline: 32680607]
2. Eriksson L, Sundman E, Olsson R, Nilsson L, Witt H, Ekberg O, et al. Functional assessment of the pharynx at rest and during swallowing in partially paralyzed humans: simultaneous videomanometry and mechanomyography of awake human volunteers. Anesthesiology 1997 Nov;87(5):1035-1043 [FREE Full text] [doi: 10.1097/00000542-199711000-00005] [Medline: 9366453]
3. Sundman E, Witt H, Olsson R, Ekberg O, Kuylenstierna R, Eriksson LI. The incidence and mechanisms of pharyngeal and upper esophageal dysfunction in partially paralyzed humans: pharyngeal videoradiography and simultaneous manometry after atracurium. Anesthesiology 2000 Apr;92(4):977-984 [FREE Full text] [doi: 10.1097/00000542-200004000-00014] [Medline: 10754616]
4. Eriksson L, Sato M, Severinghaus JW. Effect of a vecuronium-induced partial neuromuscular block on hypoxic ventilatory response. Anesthesiology 1993 Apr;78(4):693-699 [FREE Full text] [doi: 10.1097/00000542-199304000-00012] [Medline: 8096684]
5. Berg H, Roed J, Viby-Mogensen J, Mortensen CR, Engbaek J, Skovgaard LT, et al. Residual neuromuscular block is a risk factor for postoperative pulmonary complications. A prospective, randomised, and blinded study of postoperative pulmonary complications after atracurium, vecuronium and pancuronium. Acta Anaesthesiol Scand 1997 Oct;41(9):1095-1103 [FREE Full text] [doi: 10.1111/j.1399-6576.1997.tb04851.x] [Medline: 9366929]
6. Brull S, Kopman AF. Current Status of Neuromuscular Reversal and Monitoring: Challenges and Opportunities. Anesthesiology 2017 Jan;126(1):173-190 [FREE Full text] [doi: 10.1097/ALN.0000000000001409] [Medline: 27820709]
7. Naguib M, Brull S, Kopman A, Hunter J, Fülesdi B, Arkes H, et al. Consensus Statement on Perioperative Use of Neuromuscular Monitoring. Anesthesia & Analgesia 2018;127(1):71-80 [FREE Full text] [doi: 10.1213/ane.0000000000002670]
8. Baillard C, Clec'h C, Catineau J, Salhi F, Gehan G, Cupa M, et al. Postoperative residual neuromuscular block: a survey of management. Br J Anaesth 2005 Nov;95(5):622-626 [FREE Full text] [doi: 10.1093/bja/aei240] [Medline: 16183681]
9. Murphy G, Szokol JW, Marymont JH, Greenberg SB, Avram MJ, Vender JS. Residual neuromuscular blockade and critical respiratory events in the postanesthesia care unit. Anesth Analg 2008 Jul;107(1):130-137 [FREE Full text] [doi: 10.1213/ane.0b013e31816d1268] [Medline: 18635478]

10. Murphy G, Szokol JW, Avram MJ, Greenberg SB, Marymont JH, Vender JS, et al. Intraoperative acceleromyography monitoring reduces symptoms of muscle weakness and improves quality of recovery in the early postoperative period. Anesthesiology 2011 Nov;115(5):946-954 [FREE Full text] [doi: 10.1097/ALN.0b013e3182342840] [Medline: 21946094]
11. Harman A, Tung A, Fox C, Lien C. Heuristics, Overconfidence, and Experience. Anesthesia & Analgesia 2019;128(6):1057-1059 [FREE Full text] [doi: 10.1213/ane.0000000000003965]
12. Viby-Mogensen J, Jensen E, Werner M, Nielsen HK. Measurement of acceleration: a new method of monitoring neuromuscular function. Acta Anaesthesiol Scand 1988 Jan;32(1):45-48 [FREE Full text] [doi: 10.1111/j.1399-6576.1988.tb02686.x] [Medline: 2830760]
13. Kopman A, Klewicka MM, Neuman GG. The relationship between acceleromyographic train-of-four fade and single twitch depression. Anesthesiology 2002 Mar;96(3):583-587 [FREE Full text] [doi: 10.1097/00000542-200203000-00012] [Medline: 11873031]
14. Capron F, Alla F, Hottier C, Meistelman C, Fuchs-Buder T. Can acceleromyography detect low levels of residual paralysis? A probability approach to detect a mechanomyographic train-of-four ratio of 0.9. Anesthesiology 2004 May;100(5):1119-1124 [FREE Full text] [doi: 10.1097/00000542-200405000-00013] [Medline: 15114208]
15. Todd M, Hindman B, King B. The Implementation of Quantitative Electromyographic Neuromuscular Monitoring in an Academic Anesthesia Department. Anesthesia & Analgesia 2014;119(2):323-331 [FREE Full text] [doi: 10.1213/ane.0000000000000261]
16. Carvalho H, Verdonck M, Berghmans J, Poelaert J. Development and validation of an android-based application for anaesthesia neuromuscular monitoring. J Clin Monit Comput 2019 Oct;33(5):863-870 [FREE Full text] [doi: 10.1007/s10877-018-0224-2] [Medline: 30446866]
17. Aggarwal C. Outlier analysis. In: Data Mining. Berlin, Germany: Springer; 2015:237-263.
18. Chandola V, Banerjee A, Kumar V. Anomaly detection. ACM Comput. Surv 2009 Jul;41(3):1-58 [FREE Full text] [doi: 10.1145/1541880.1541882]
19. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
20. Cordova Plugin Device-Motion. URL: https://github.com/apache/cordova-plugin-device-motion [accessed 2021-05-25]
21. Krueger E, Scheeren E, Nogueira-Neto G, Button V, Nohama P. Advances and perspectives of mechanomyography. Rev. Bras. Eng. Bioméd 2014 Dec;30(4):384-401 [FREE Full text] [doi: 10.1590/1517-3151.0541]
22. Scikit-Learn. URL: https://scikit-learn.org/ [accessed 2021-05-24]
23. Python Programming Language. URL: https://www.python.org/ [accessed 2021-05-24]
24. Naguib M, Kopman AF, Lien CA, Hunter JM, Lopez A, Brull SJ. A Survey of Current Management of Neuromuscular Block in the United States and Europe. Anesthesia & Analgesia 2010;111(1):110-119 [FREE Full text] [doi: 10.1213/ane.0b013e3181c07428]
25. Dahaba A, Rehak PH, List WF. Assessment of accelerography with the TOF-GUARD: a comparison with electromyography. Eur J Anaesthesiol 1997 Nov;14(6):623-629 [FREE Full text] [doi: 10.1017/s026502159700104x] [Medline: 9466099]
26. Nakata Y, Goto T, Saito H, Ichinose F, Uezono S, Suwa K, et al. Comparison of Acceleromyography and Electromyography in Vecuronium-induced Neuromuscular Blockade with Xenon or Sevoflurane Anesthesia. Journal of Clinical Anesthesia 1998 May;10(3):200-203 [FREE Full text] [doi: 10.1016/s0952-8180(98)00006-3]
27. Jung W, Hwang M, Won YJ, Lim BG, Kong MH, Lee IO. Comparison of clinical validation of acceleromyography and electromyography in children who were administered rocuronium during general anesthesia: a prospective double-blinded randomized study. Korean J Anesthesiol 2016 Feb;69(1):21-26 [FREE Full text] [doi: 10.4097/kjae.2016.69.1.21] [Medline: 26885297]
28. Kopman A, Chin W, Cyriac J. Acceleromyography vs. electromyography: an ipsilateral comparison of the indirectly evoked neuromuscular response to train-of-four stimulation. Acta Anaesthesiol Scand 2005 Mar;49(3):316-322 [FREE Full text] [doi: 10.1111/j.1399-6576.2005.00643.x] [Medline: 15752395]
29. Liang S, Stewart P, Phillips S. An Ipsilateral Comparison of Acceleromyography and Electromyography During Recovery from Nondepolarizing Neuromuscular Block Under General Anesthesia in Humans. Anesthesia & Analgesia 2013;117(2):373-379 [FREE Full text] [doi: 10.1213/ane.0b013e3182937fc4]
30. Etemadi M, Hogue CW. Preventing Intraoperative Hypotension: Artificial Intelligence versus Augmented Intelligence? Anesthesiology 2020 Dec;133(6):1170-1172 [FREE Full text] [doi: 10.1097/ALN.0000000000003561] [Medline: 33380745]
31. Bowdle A, Bussey L, Michaelsen K, Jelacic S, Nair B, Togashi K, et al. A comparison of a prototype electromyograph vs. a mechanomyograph and an acceleromyograph for assessment of neuromuscular blockade. Anaesthesia 2020 Feb;75(2):187-195 [FREE Full text] [doi: 10.1111/anae.14872] [Medline: 31617199]

Abbreviations

AMG: acceleromyography
AUC: area under the curve

CSLR: cost-sensitive logistic regression
EMG: electromyography
ROC: receiver operator characteristics
TOF: train-of-four
TOFR: train-of-four ratio

Edited by R Kukafka; submitted 23.11.20; peer-reviewed by D Schwarz, L Guo; comments to author 09.03.21; revised version received 18.03.21; accepted 04.05.21; published 21.06.21.

Please cite as:
Verdonck M, Carvalho H, Berghmans J, Forget P, Poelaert J
Exploratory Outlier Detection for Acceleromyographic Neuromuscular Monitoring: Machine Learning Approach
J Med Internet Res 2021;23(6):e25913
URL: https://www.jmir.org/2021/6/e25913/
doi:10.2196/25913
PMID:34152273

Original Paper

Semantic Linkages of Obsessions From an International Obsessive-Compulsive Disorder Mobile App Data Set: Big Data Analytics Study

Jamie D Feusner[1,2*], MD; Reza Mohideen[3*]; Stephen Smith[3], BA; Ilyas Patanam[3], BSc; Anil Vaitla[3], BSc; Christopher Lam[3], BSc; Michelle Massi[4], BSc, MA; Alex Leow[5], MD, PhD

[1]Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA, United States
[2]Centre for Addiction and Mental Health, Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[3]NOCD, LLC, Chicago, IL, United States
[4]Anxiety Therapy LA, Inc, Los Angeles, CA, United States
[5]Department of Psychiatry, University of Illinois College of Medicine, Chicago, IL, United States
*these authors contributed equally

Corresponding Author:
Jamie D Feusner, MD
Semel Institute for Neuroscience and Human Behavior
University of California Los Angeles
300 UCLA Medical Plaza
Suite 2200
Los Angeles, CA, 90095
United States
Phone: 1 3102064951
Email: jfeusner@mednet.ucla.edu

Abstract

Background: Obsessive-compulsive disorder (OCD) is characterized by recurrent intrusive thoughts, urges, or images (obsessions) and repetitive physical or mental behaviors (compulsions). Previous factor analytic and clustering studies suggest the presence of three or four subtypes of OCD symptoms. However, these studies have relied on predefined symptom checklists, which are limited in breadth and may be biased toward researchers' previous conceptualizations of OCD.

Objective: In this study, we examine a large data set of freely reported obsession symptoms obtained from an OCD mobile app as an alternative to uncovering potential OCD subtypes. From this, we examine data-driven clusters of obsessions based on their latent semantic relationships in the English language using word embeddings.

Methods: We extracted free-text entry words describing obsessions in a large sample of users of a mobile app, *NOCD*. Semantic vector space modeling was applied using the Global Vectors for Word Representation algorithm. A domain-specific extension, *Mittens*, was also applied to enhance the corpus with OCD-specific words. The resulting representations provided linear substructures of the word vector in a 100-dimensional space. We applied principal component analysis to the 100-dimensional vector representation of the most frequent words, followed by k-means clustering to obtain clusters of related words.

Results: We obtained 7001 unique words representing obsessions from 25,369 individuals. Heuristics for determining the optimal number of clusters pointed to a three-cluster solution for grouping subtypes of OCD. The first had themes relating to relationship and just-right; the second had themes relating to doubt and checking; and the third had themes relating to contamination, somatic, physical harm, and sexual harm. All three clusters showed close semantic relationships with each other in the central area of convergence, with themes relating to harm. An equal-sized split-sample analysis across individuals and a split-sample analysis over time both showed overall stable cluster solutions. Words in the third cluster were the most frequently occurring words, followed by words in the first cluster.

Conclusions: The clustering of naturally acquired obsessional words resulted in three major groupings of semantic themes, which partially overlapped with predefined checklists from previous studies. Furthermore, the closeness of the overall embedded relationships across clusters and their central convergence on harm suggests that, at least at the level of self-reported obsessional thoughts, most obsessions have close semantic relationships. Harm to self or others may be an underlying organizing theme across many obsessions. Notably, *relationship*-themed words, not previously included in factor-analytic studies, clustered with *just-right* words. These novel insights have potential implications for understanding how an apparent multitude of obsessional symptoms are connected by underlying themes. This observation could aid exposure-based treatment approaches and could be used as a conceptual framework for future research.

(J Med Internet Res 2021;23(6):e25482) doi:10.2196/25482

KEYWORDS

OCD; natural language processing; clinical subtypes; semantic; word embedding; clustering

Introduction

Background

Obsessive-compulsive disorder (OCD) is characterized by recurrent and persistent thoughts, urges, or images that are experienced as intrusive and inappropriate (obsessions) and that cause marked anxiety or distress and/or repetitive behaviors (compulsions) [1]. OCD has a lifetime prevalence of approximately 2%-2.3% worldwide [2,3] and is associated with functional impairment, poor quality of life, and increased use of health care services [4].

OCD symptoms can manifest in a variety of seemingly disparate ways [1]. Obsessions can be, for example, fears of contamination from one's immediate environment, ego-dystonic thoughts that one might harm someone else in a violent or sexual manner despite not having a desire or intent to do so, believing that one has done something blasphemous, excessive doubts or concerns about one's partner in a relationship, or having a difficult-to-describe need to feel that an object is placed *just right* in their environment. In response to these obsessions, patients might engage in different types of compulsive behaviors that could include, for example, handwashing, checking, praying, arranging items, mental self-reassurance, or avoiding situations that trigger obsessive thoughts.

This range of different symptoms has contributed to the notion that OCD may be a heterogeneous condition characterized by different *subtypes* [5,6]. If so, subtypes might have important implications for understanding the potentially different underlying neurobiology and may have clinical implications, such as different effective treatment approaches for different subtypes.

To date, multiple factor-analytic studies have been conducted on OCD to understand the aggregations of symptoms and establish different potential subtypes [7,8]. Most of these studies have focused on symptoms elicited from clinical interviews rather than underlying pathophysiological mechanisms or processes as the basis for subtyping [7]. Symptom category–based factor-analytic studies have demonstrated evidence for four dimensions or subgroups of obsessions: contamination, harming, symmetry or order, and hoarding (hoarding has since been reconceptualized in the Diagnostic and Statistical Manual-5 as a separate disorder) [9,10] or contamination or somatic, aggressive, sexual, or religious, symmetry or order, and hoarding [11,12]. Two studies found evidence for five latent structures: contamination, harming, symmetry or order, hoarding, and religious or sexual [13] or sexual or somatic [14]. Cluster analyses have also been applied to identify subgroups of OCD symptoms, with some finding similar results as the factor-analytic studies [15,16]; however, one analysis found evidence of additional obsessional symptom clusters of sexual or somatic and contamination or harming [17].

A meta-analysis of 21 studies and 5124 participants found four factors using category-level data: cleaning and contamination, aggressive, sexual, or somatic obsessions (*forbidden thoughts*), symmetry, and hoarding [8]. Using item-level data revealed a five-factor solution: symmetry, aggression or sexual or religious, contamination, aggression or checking, and somatic.

These studies and most approaches to assess OCD symptoms to date have relied on assessing the presence of symptoms by selection from the predefined sets of obsessions and compulsions, most commonly from the clinician-administered Yale-Brown Obsessive-Compulsive Scale Symptom Checklist (YBOCS-SC) [18]. The individual items on the YBOCS-SC and the clustering of these into 13 categories was originally established by clinical consensus. However, using a predefined list of symptoms may introduce bias as it could lead to patients and/or clinicians *fitting* the symptoms into the terms on the checklist, which are further arranged in predefined categories.

Apart from handwashing, the YBOCS-SC was shown in one study to have poor convergent validity with self-report measures [17]. However, this may be attributed to the incomplete representation of symptoms in the other self-report measures to which the YBOCS-SC was compared, which also has the limitation of having predefined checklists. For instance, relationship-related obsessions [19-24] pertaining to obsessions regarding the suitability of the relationship or the relationship partner are not in the YBOCS-SC. Another study found an incomplete correspondence between the YBOCS-SC and self-report measures [25]. Furthermore, its internal consistency reliability was low for symmetry or ordering and sexual or religious symptoms; however, it was adequate for contamination or handwashing and aggression or checking [26].

A potentially less biased approach to assess and classify the types of OCD symptoms could come from patients' free responses rather than relying on checklists. However, using free responses to understand the relationships among these symptoms to determine if categories or factors exist poses a challenge because of the vast number of different possible responses. Very large samples of responses, from large numbers of individuals, are likely required to identify stable patterns because multiple repeated terms, or terms representing similar semantic themes of symptoms, would be needed to separate signals from noise.

Such large samples could be obtained using digitally obtained data, such as from mobile apps used by individuals with OCD [27]. Mobile digital sources of free-entry text could provide a valuable resource for uncovering naturalistic, unrestricted, spontaneous, and therefore relatively unbiased patterns of responses.

Although they are more difficult to classify as compared with checklists, the semantic relationships of freely entered words representing obsessions can be analyzed for their clustering relationships in the English language. A tool to perform this clustering is natural language processing (NLP)—a branch of artificial intelligence that attempts to bridge the gap between computers and humans using the natural language [28]. Some common use cases of NLP include language translation apps such as Google Translate and computerized personal assistants such as Siri and Google Assistant. Each of these apps rely on NLP to decipher and understand human languages to complete a task. Approaches using NLP could provide a better understanding of latent semantic themes that could form the organizing relationships of groupings of protean individual obsessional examples. As a result, we apply NLP to characterize obsessions in this first-of-its-kind study.

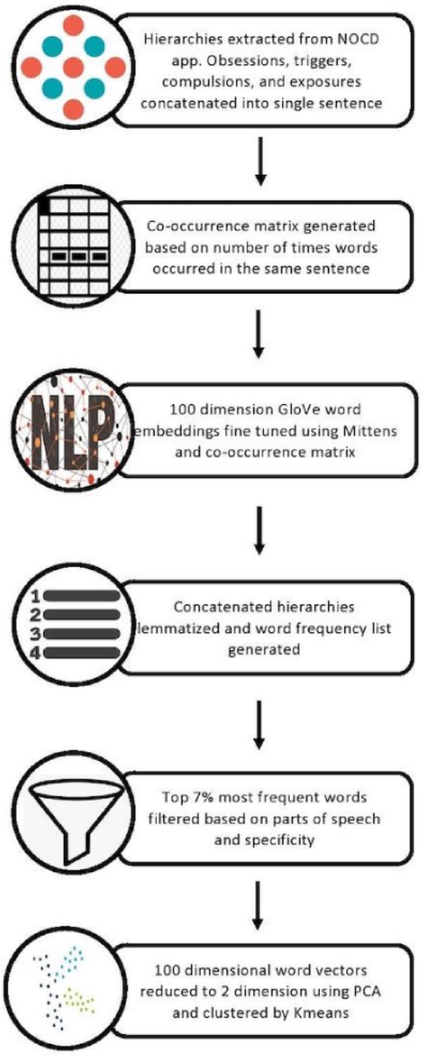
Objectives

To explore this approach, we obtained free-entry data for obsessions from a mobile health treatment platform developed by NOCD [29]. The NOCD app, among other functions, provides users with a platform for setting up customizable exposure and response prevention (a form of cognitive behavioral therapy for OCD) exercises. The primary objective of this study is to determine the number, types, and relationships of semantically organized clusters of obsessions in a data-driven manner. To this end, we applied an NLP technique called word embedding to a large data set of words from a large sample of individuals using the NOCD app.

Methods

An overview of the data extraction and following processing steps are shown in Figure 1.

Figure 1. Data extraction and processing steps. GloVe: Global Vectors for Word Representation; PCA: principal component analysis.



Data Extraction

We extracted data from freely entered words describing obsessions and their corresponding triggers, exposures, and compulsions by users of the iOS version of the mobile app NOCD who self-identified as having OCD (NOCD iOS app version 2.0.7-2.0.96). This app includes two features where users can input their current obsessions, namely, the “SOS” feature and the hierarchy, which can then be used for planned exposure exercises. Obsessions that had been inputted by users were assigned an anonymous user ID number, and all data were deidentified. NOCD users are required to be at least 17 years of age, but to preserve privacy, no specific demographic information was collected. Data were collected between March 22, 2018, and July 9, 2020.

Data Preprocessing

Obsessions and descriptions of obsessions from their associated triggers, exposures, and compulsions were grouped together into a singular phrase. In addition, phrases were cleaned to remove punctuation, special characters, and spelling errors.

Generating Word Embeddings—Global Vectors for Word Representation Algorithm and Mittens

A co-occurrence matrix of size N×N is generated by iterating over the aforementioned phrases and incrementing (i,j), where N is the number of unique words, and i and j correspond to two words that appear in the same phrase. Pretrained 100-dimension Global Vectors for Word Representation (GloVe) algorithm [30] word-embedding vectors were used in conjunction with the co-occurrence matrix as inputs to Mittens [31], which is an extension of GloVe for learning domain-specialized representations.

Thus, GloVe vectors, based on global word-to-word co-occurrence statistics from a 6 billion–word corpus, were fine-tuned based on the localized co-occurrence matrix. This provides a specialized representation of words and their relationships from the OCD-specific lexicon synthesized with pretrained representations from the general lexicon.

Data Processing

To identify the root forms of words that may be inflections of each other, we performed lemmatization using Stanford CoreNLP—natural language software [32]. The entries were first parsed into single words. These words were assigned a frequency as to how often they occurred across all individuals’ entries and then sorted by the frequency of occurrence. The top percentage of most frequently occurring words was chosen; the top 7% was selected as it captured the most commonly occurring words balanced with the ability to visualize the cluster graphs.

These words were then filtered based on the parts of speech: adverbs, modals, third-person singular present verbs, gerunds, past participle verbs, *to*, prepositions, subordinating conjunctions, and personal pronouns. In addition, a clinician (JDF) reviewed the filtered list to further remove nonclinically relevant words, resulting in a total of 430 words.

Clustering and Data Analysis

To visualize the 100-dimension latent semantic relationships of words in a 2D space, we applied principal component analysis to the 100-dimensional vector representation of the most frequent words and plotted the first two principal components. Furthermore, to identify clusters of related words in a data-driven manner, we performed k-means clustering. The optimal number of clusters was determined using the following heuristics: the silhouette coefficient [33], Elbow Method [34], Calinski-Harabasz Index [35], and the Davies-Bouldin index [36].

Once the optimal cluster number was determined, we compared the relative frequencies among the clusters of (1) unique obsessional words and (2) total obsessional words using the chi-square test.

Results

Characterization of the Data

We obtained 7001 unique words representing obsessions from 25,369 individuals aged 17 years and older, self-identified as having OCD across 108 countries. Most individuals were from the United States (18,315) with an additional 1335 North American entries from Canada and Mexico, 4134 from Europe, 557 from Asia, 100 from Africa, 137 from South America, and 791 from Australia and New Zealand.

In total, 94.99% (24,100/25,369) of users contributed no more than five obsessions each to the data set. Most users—16,988 (the mode)—contributed only one obsession; 4311 contributed two obsessions, 1861 contributed three obsessions, 854 contributed four obsessions, and 476 contributed five obsessions. There were two extreme outliers that contributed 120 and 174 obsessions. We removed these data before the analysis. In summary, although users could enter multiple words, given the large total number of obsessional words that were mapped and the fact that most users only entered one or two words, the overall results were unlikely to be biased by single users (Multimedia Appendix 1, Figure S5, shows a histogram of the number of word entries per user).

Determining Cluster Size

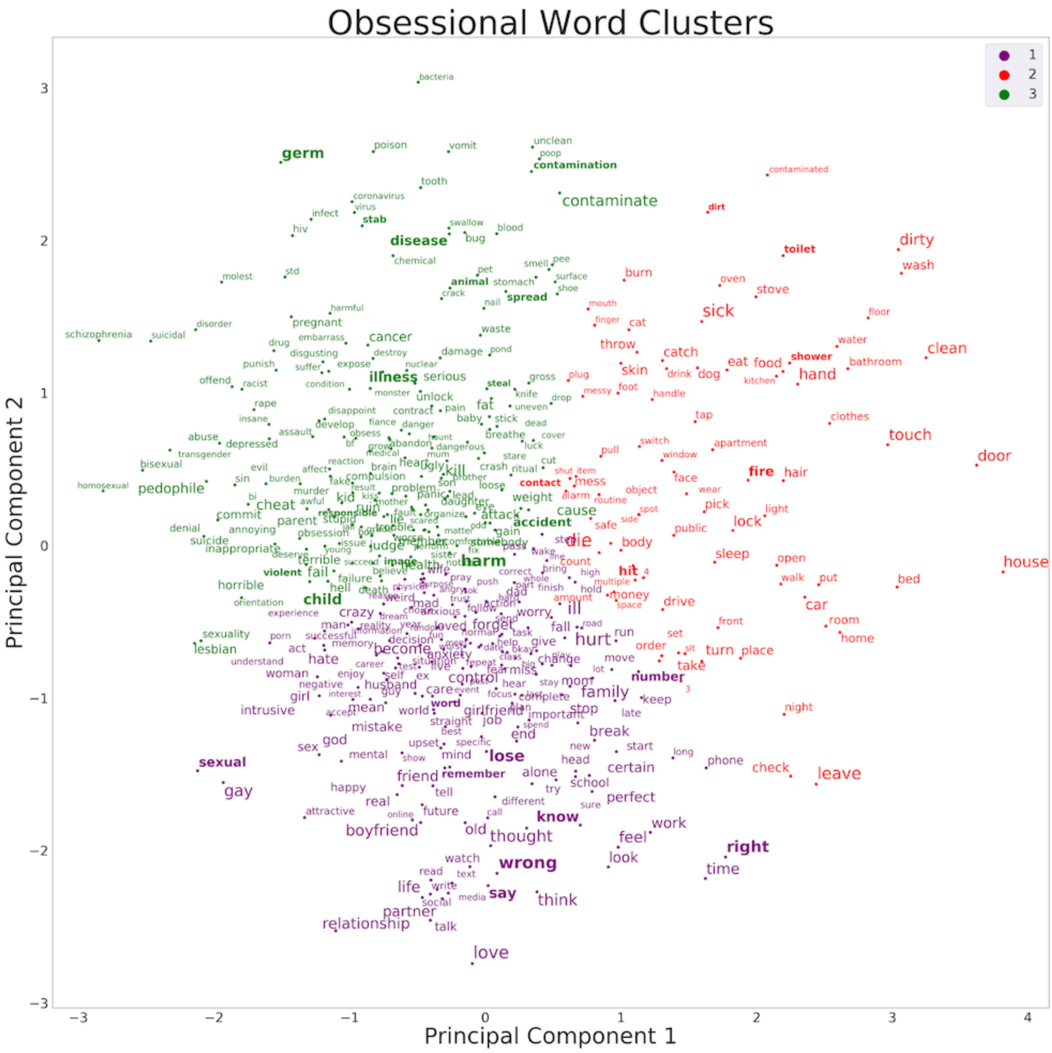
The silhouette coefficient yielded optimal clusters with sizes of 3 and 5. The Elbow Method, Calinski-Harabasz index, and Davies-Bouldin index all pointed to k=3 as the optimal number of clusters (Multimedia Appendix 1, Figure S5).

Relationships of Obsessional Words With Canonical OCD Symptom Factors

Next, we determined how the clusters of obsessions from our data-driven methods using freely reported obsessional words compared with previous factor-analytic and clustering studies that characterized symptom groupings based on rating scale checklists. To do this, we qualitatively examined which cluster appeared as OCD-specific words from the YBOCS-SC. We examined this for the optimal cluster solution of k=3 (Figure 2) [16] as well as for k=2, k=4, and k=5 clusters (Figure 3). Across all cluster solutions, there was a large, dense, central grouping with principal themes relating to harming of self or others (eg, *harm*, *accident*, and *hit*).

Furthermore, across k=3, k=4, and k=5 cluster solutions, the densest regions of each cluster were at a central convergence of the clusters, thereby suggesting closely grouped embedded relationships both within and across clusters (including on the borders between these clusters). Furthermore, cluster *3* in the k=3 solution that contained contamination and somatic-themed words (eg, *contamination*, *germ*, *disease*, and *illness*) and physical- and sexual-related harm words (eg, *harm*, *accident*, *child*, and *sexual*) split from each other in the k=4 solutions, suggesting that although these themes are related, distinctness is evident at the next level of separation.

Figure 2. Frequently occurring obsessional words and their clustering, based on semantic relationships. The word embedding was trained on the entirety of the data set and clustered using k-means with k=3 clusters. The font is scaled according to the frequency of occurrence of each word. For reference, bolded words are those that also appear in the Yale-Brown Obsessive-Compulsive Scale Symptom Checklist.



Split-Sample Repeat of the Clustering Analysis

We repeated the analysis in an equal-sized, nonchronological overlapping sample to control for any possible influences of minor changes in the NOCD app user interface that occurred during the period of data collection. The first sample was from March 22, 2018, to August 14, 2019 (11:18:04 PM), and the second sample was from August 14, 2019 (11:42:12 PM), to July 9, 2020, and each sample included 22,749 and 22,750 words, respectively.

The two aforementioned samples demonstrated similar clustering patterns, suggesting that the results were stable over time. To quantify this observation, we compared the 2D embeddings of the two groups by calculating the distance from a word to all other words on the graph. We performed this task for all words, creating a matrix of distances. Matrices for the two groups were compared using the Pearson correlation coefficient. Most correlations were above r=0.90, thereby demonstrating a high consistency of results across separate periods (Multimedia Appendix 1, Figure S14).

As the two abovementioned samples included some of the same individuals who entered obsessions in both the first and second periods, we additionally repeated the analysis in two nonoverlapping equal-sized user samples. The first sample contained 12,684 users with 22,510 obsessions and the second sample contained 12,685 users with 22,989 obsessions. By performing the same comparison of word-to-word matrices as for the chronological split sample, we found that the two split-by-users samples demonstrated very similar clustering patterns with the majority of correlations $r \geq 0.90$, suggesting that the results are reliable over subsets of users.

Relative Frequency of Obsessional Words by Cluster

Comparing the three clusters on the total number of obsessional words, we observed 169 words in cluster 1 (just-right and relationship themes), 86 words in cluster 2 (doubt or checking themes), and 174 words in cluster 3 (contamination, somatic, or harm themes). All three clusters were significantly different from each other ($\chi^2=34.2$; $P<.001$). Furthermore, cluster 1 was significantly different from cluster 2 ($\chi^2=27.4$; $P<.001$) and cluster 3 ($\chi^2=11.4$; $P<.001$). Cluster 2 was significantly different from cluster 3 ($\chi^2=29.4$; $P<.001$). Thus, contamination, somatic, or harm words were more frequent than the just-right and relationship words and the doubt or checking words, and the just-right or relationship words were more frequent than the doubt or checking words.

Discussion

Principal Findings

In this large data set, we applied a data-driven approach and the NLP technique of word embedding to understand commonly occurring semantic themes of obsessions, freely reported by individuals using an OCD app. The optimal number of independent clusters that represents the relationships of the most frequently occurring obsessions was 3. Notably, the embedding patterns revealed that most obsessional words were closely grouped within and across clusters, including on the borders between clusters. Moreover, the densest region of each cluster is at a central convergence of the three, with principal themes relating to harm. This suggests that, at the level of self-reported obsessional thoughts, most obsessions have close semantic links with each other. Thus, although unique obsessions are protean, many examples, even across cluster *subtypes*, may actually have underlying latent relationships with each other.

Observations of Semantic Themes of Clusters

To relate these findings to previous studies of OCD subtypes, we chose descriptive labels for the clusters from words from the widely used YBOCS-SC from six of the eight obsession categories: aggressive, sexual, religious, somatic, symmetry, and contamination (excluding hoarding and miscellaneous). Using these labels, the optimal cluster solution of three resulted in contamination-themed and physical- and sexual harm–themed obsessions occurring in the same cluster (cluster 3) and relationship and just-right themes occurring in the same cluster (cluster 1; Figure 2). Cluster 2 included many doubt-related obsessions typically associated with checking compulsions and a subset of contamination-related obsessions (eg, toilet, shower, sick, or dirty).

A notable overall observation is that the clustering algorithm produced topological patterns demonstrating a large but diffuse central cluster of harm-themed words, in addition to more diffuse and distanced words that represent contamination themes and doubt-related obsessions typically associated with checking compulsions (eg, fire, house, door, car, or lock). This is most readily apparent at the lowest level of clustering (k=2; Figure 3). As the number of clusters increases from k=3 to k=4 to k=5, some contamination-themed words joined with somatic-related themes (cluster 3 for k=3; Figure 3), which subsequently split into two different clusters (clusters 3 and 4 for k=4; Figure 3). At k=5, the original harm cluster from k=2 split into three clusters (clusters 1, 3, and 4), comprising words with themes related to sexual, sexual orientation, and relationships (cluster 1); *just-right* (cluster 3); and physical harm and a subset of doubt or checking (cluster 4).

This latter observation that a subset of doubt or checking words split off into two separate clusters at this level is consistent with a previous factor-analytic study finding that *aggressive* obsessions and checking compulsions tend to show instability on which factors they load on [37]. This progressive splitting of groups of obsessions as the cluster number increases provides unique insights into how closely some obsessional themes are related to each other based on their semantic relationships embedded in a 100-dimensional space.

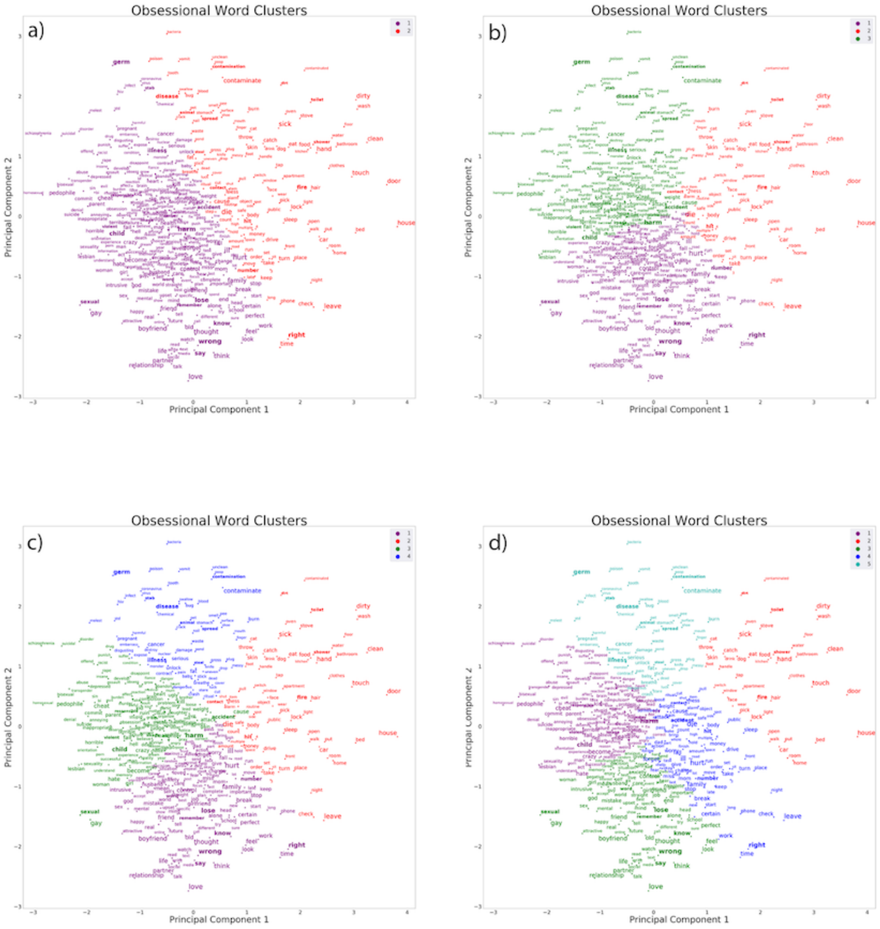
These results are largely consistent with the findings of previous factor-analytic studies. A previous meta-analysis [8] of factor-level results similarly suggested a three-factor solution (not including hoarding symptoms); however, the item-level results pointed to five factors. There are several possible reasons why results differ, at least partially, from previous studies. First, this study represents an obsession from a much larger sample size: 25,369 individuals. The previous meta-analysis included a total of 5124 individuals; yet, individual study sizes ranged from 45 to 615. Thus, this study might have sampled a broader population.

Furthermore, as described in the Introduction section, previous factor-analytic and clustering studies used data collected with existing scales that include predefined checklists and categories, often from the YBOCS-SC. This could have the effect of creating a conceptual framework into which patients (and clinicians and researchers) may *shoe-horn* their experiences and symptoms. In addition, if a patient has an obsessional preoccupation that is not listed on the scale, then the patient, clinician, or researcher may not identify it as an OCD symptom. Freely entered text, as the NOCD app allows, mitigates this limitation. Notwithstanding, many individuals using the app may have already been exposed to commonly used lexicons regarding OCD subtypes, from clinicians, educational material, the media, or even the NOCD app itself because the online *community* forum mentions OCD subtypes.

Additional factors might account for why results partially diverge from previous factor-analytic and clustering studies. Some patients may feel embarrassed, ashamed, or even fearful of the consequences of divulging certain *taboo* obsessional thoughts, such as blasphemous, pedophilic, other sexual, or violent themes in front of a clinician or researcher. Such individuals may find it easier to enter these, on their own, in an app. In addition, in the online community forum in NOCD, people can find others who share similar thoughts. Once they have identified that what they experience is likely an OCD symptom, because it is shared by others, this could facilitate them entering the obsessional words related to these themes in the *hierarchy* part of the app. In sum, this approach of mapping the embedded latent relationships of obsessions partially recapitulates factor-analytic studies' findings; yet, the differences might reflect a more unconstrained capturing of naturally occurring obsessions.

The split-sample test over time and across unique individuals demonstrated overall stable results. This stability across time suggests that the semantic themes of obsessional words entered by users are mostly invariant to periodic updates and fixes that happened with NOCD, as is standard with the most widely used mobile apps. In addition, early adopters of apps in general may have certain characteristics that differ from later adopters [38]. However, the stability across the split sample demonstrates that the results hold across these potentially different sets of individuals.

Figure 3. Frequently occurring obsessional words and their clustering, based on semantic relationships. The word embedding was trained on the entirety of the data set and clustered with k values of 2, 3, 4, and 5. The font is scaled according to the frequency of occurrence of each word. For reference, bolded words are those that also appear in the Yale-Brown Obsessive-Compulsive Scale Symptom Checklist. Observations of how the bolded words change clusters as k values change provide valuable insight into the similarities of OCD subtypes.



Relative Frequencies of Obsessions Across Clusters

Earlier studies that categorized patients with OCD according to primary compulsions observed that contamination obsessions were the most common [39]. A subsequent study found that cleaning compulsions (therefore most often associated with contamination fears) were almost twice as common as checking compulsions (most often associated with harm or doubt obsessions) and more than four times as common as obsessions without physical compulsions in people with OCD seeking treatment [40]. A community assessment of OCD from the National Comorbidity Survey Replication included 2073 people and administered the YBOCS [2]. The results suggested that obsessions leading to checking (typically caused by obsessional fear of harm or doubt) are more common than contamination fears. The latter finding could have included the types of harm obsessions that were found to be highly represented in this study's data set (clusters 2 and 3 in Figure 3), although it is unclear from the data collected in that study.

In this study's sample, obsessions with harm themes were represented in a much higher proportion than in previous studies. Although harm-, contamination-, and somatic-themed words were clustered together in the k=3 grouping, the harm and contamination or somatic themes split and harm themes were nearly twice as common (n=138) as contamination or somatic words (n=70) at k=4. As most individuals in the current sample contributed only one obsession, this might represent their primary symptom, although this was not specifically ascertained. One interpretation of this finding is that this could represent a truer reflection of the relative proportions of these obsessional themes in the larger population of those with OCD, particularly because this study had a sample size that was an order of magnitude larger than the National Comorbidity Survey

Replication. Furthermore, previous studies may not have found such relative proportions because of the limited thematic content of the YBOCS-SC or other instruments used to collect these data (as mentioned earlier). In addition, as mentioned earlier, *taboo* themes often relate to harming others and might appear more frequently in this data set obtained via relatively anonymous entry as compared with other data sets from population-based studies that involved researchers directly querying participants about subtypes.

An alternative interpretation is that there may be an overrepresentation of individuals who use the NOCD app who have harm obsessions. If this is the case, then it could be related to the stigmatization and *taboo* nature of these themes. Individuals with unwanted obsessional thoughts that are taboo in most societies and cultures, such as sexual thoughts about children, incest, physically harming someone, or blasphemous religious thoughts, may not readily share them with clinicians or researchers. However, they may feel more willing to share these in an app, which is a more anonymous experience. In addition, the NOCD app includes an online community forum in which other individuals share topics such as obsessional thoughts. This may help individuals not only realize that these thoughts could be related to OCD (they may not have found exact examples of their particular obsession elsewhere) but also the stigmatization of sharing them and entering them in the app could be reduced.

Another possibility that could lead to the overrepresentation of individuals with *harm* themes among NOCD users is that they might not have received effective treatment elsewhere and therefore gravitated toward NOCD as an alternative to try to find help. A study by primary care physicians demonstrated that OCD was misdiagnosed as either another psychiatric or psychological condition or no diagnosis 50.5% of the time [41]. Furthermore, in that study, obsessions related to homosexuality, aggression, and pedophilia were misdiagnosed ≥70% of the time. A similar study by clinical psychologists found that OCD was misdiagnosed 38.9% of the time and obsessions about the taboo thoughts of homosexuality, sexual obsessions about children, aggressive obsessions, and religious obsessions were more frequently misdiagnosed than contamination obsessions [42]. Even if diagnosed correctly, clinicians might find these types of symptoms difficult to understand and/or difficult to treat with exposure and response prevention because the themes are outside of the well-known *textbook* examples of contamination fears and checking compulsions, many of which have mental rather than physical compulsions or primarily engage in avoidance behaviors.

Relationship-Themed Obsessions

Obsessions related to relationships have recently received attention in clinical and research settings [19-24], spawning the term "relationship OCD." Although the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* [1] does not explicitly describe "relationship obsessions," it provides an example under Other Specified Obsessive-Compulsive and Related Disorder, "Obsessional jealousy...a nondelusional preoccupation with a partner's perceived infidelity." Other partner-centered obsessional themes include obsessions about a partner's flaws, such as intelligence, social aptitude, and morality [21]. Additional themes are centered on the relationship itself and can include obsessional doubting about whether one's relationship is a *good* or *ideal* relationship. In addition to having obsessional *thoughts* about relationships, this can also include intrusive images, urges, or *not right* feelings.

To the best of our knowledge, this study is the first to characterize relationship-themed obsessions using clustering or factor-analytic approaches. Many such relationship themes appear in the same cluster as "just-right" obsessions, sometimes referred to as "not-just-right-experiences" [43]. This suggests the possibility that a driving force for many with relationship-themed obsessions could be that something about it does not *feel right*, rather than, for example, other catastrophic or otherwise negative consequences related to the relationship itself or one's partner.

Potential Clinical Implications

There are several potentially important insights that this study provides to the semantic relationships of obsessional thoughts. These have potential clinical assessment and treatment implications for understanding how an apparent multitude of *surface* obsessional symptoms are connected by underlying themes. This could assist clinicians during the assessment phase to facilitate a more focused and personalized inquiry about the presence of additional obsessions that may be semantically related to those reported. This could be helpful, particularly because patients with OCD are sometimes unaware that a particular thought is an obsession. Otherwise, it is difficult and time consuming for a clinician to guess what other obsessions they might have or to present patients with an extremely long and unfocused list of obsessions to choose from.

The findings from this study could also aid in planning exposure-based treatment approaches, for example, therapists could potentially map their patient's primary obsessional themes to understand what are nearest-neighbor themes that might tap into a yet-unexplored core obsessional fear. A specific example of the potential utility of these findings could be to help clinicians explore whether the underlying feeling or emotion associated with a relationship-themed obsession is a not-just-right experiences versus the fear of a negative consequence of the relationship.

Limitations, Strengths, and Future Directions

One of the limitations of this study is that although the data come from individuals who sought out and used therapeutic tools on an OCD app, we cannot confirm whether they met the diagnostic criteria for OCD because they did not undergo a diagnostic evaluation. It would be useful to repeat the procedures and analysis in a clinically diagnosed OCD sample; however, achieving a similar sample size would be extremely challenging. Therefore, this study's results apply to those using the NOCD app, implying that they self-identify as having OCD or an OCD-like problem and could include those with other obsessive-compulsive and related disorders, for example, body dysmorphic disorder or other disorders with prominent obsessional thinking (eg, anorexia nervosa or illness anxiety disorder). This approach is in line with alternative transdiagnostic or dimensional strategies to study psychiatric illnesses, such as the National Institute for Mental Health's Research Domain Criteria [44] framework. Another limitation is that we only examined postings in English. Although the majority of postings were in the English language, there were postings in other languages as well because NOCD is available and used worldwide. Another limitation is that we had limited demographic information because the data were strictly deidentified, limiting our ability to characterize demographic subgroups of individuals with OCD or determine whether these data represented the same gender distributions that are observed in the population of those with OCD. A further limitation is that the sample was limited to those aged 17 years and older. Whether the results can be generalized to children with OCD is unknown.

Finally, because of the variable nature of OCD and variability in the number of obsessions a user will choose to enter in the NOCD app, there was a skewed distribution of the frequency of obsessions across users (Multimedia Appendix 1, Figure S1). Thus, a small number of users theoretically could have influenced the results both at the stages of Mittens enhancement of GloVe embedding and of the comparison of frequencies of obsessions among clusters. However, to mitigate this, we removed the two extreme outliers. Furthermore, the split-sample results across equal numbers of unique users showed stable clusters and relative frequencies of obsessions, so it is unlikely that the main outcomes were influenced by a small number of individual users.

There are several important directions for future research. To understand the dynamics of semantic themes—specifically, if and how they change over time within individuals—it would be useful to obtain longitudinal data. This would allow, for example, an exploration of whether obsessional symptoms are stable in individuals or stable within semantic clusters or whether they cross over into different semantic themes. In addition, semantically defined OCD subtypes could be used as a conceptual framework to explore whether there are corresponding differences in the underlying neurobiology.

Conclusions

In this study, we analyzed the semantic relationships of obsessional words freely entered in a mobile OCD app using a large data set. This novel, data-driven method provides unique insights into the relationships of obsessional themes and identifies potential OCD subtypes. This method is distinct from traditional characterizations of phenomenology in OCD; it is not easily achieved in clinical settings or in-person research settings and circumvents the limitations and biases of pre-existing clinical and research conceptualizations of established OCD subtypes.

Acknowledgments

The authors would like to thank Nik Srinivas for his helpful comments and ideas regarding the data and analyses.

Conflicts of Interest

JDF, RM, SS, IP, AV, and CL report personal fees from NOCD Inc during the conduct of the study. AL is co-founder of KeyWise AI, and serves on the medical board of Buoy Health.

Multimedia Appendix 1

Supplemental document that contains visualizations of analysis pertaining to characterization of the data in terms of demographics and sample size, determination of cluster size, and split-sample repeat of clustering analysis.

[PDF File (Adobe PDF File), 5500 KB - jmir_v23i6e25482_app1.pdf]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). Washington, DC: American Psychiatric Publishing; 2013.
2. Ruscio AM, Stein DJ, Chiu WT, Kessler RC. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. Mol Psychiatry 2010 Jan;15(1):53-63 [FREE Full text] [doi: 10.1038/mp.2008.94] [Medline: 18725912]
3. Weissman MM, Bland RC, Canino GJ, Greenwald S, Hwu HG, Lee CK, et al. The cross national epidemiology of obsessive compulsive disorder. The Cross National Collaborative Group. J Clin Psychiatry 1994 Mar;55 Suppl:5-10. [Medline: 8077177]
4. Huppert JD, Simpson HB, Nissenson KJ, Liebowitz MR, Foa EB. Quality of life and functional impairment in obsessive-compulsive disorder: a comparison of patients with and without comorbidity, patients in remission, and healthy controls. Depress Anxiety 2009;26(1):39-45 [FREE Full text] [doi: 10.1002/da.20506] [Medline: 18800368]
5. Lewis A. Problems of obsessional illness. Proc R Soc Med 2016 Sep;29(4):325-336. [doi: 10.1177/003591573602900418]
6. Rasmussen SA, Eisen JL. Epidemiology, clinical features and genetics of obsessive-compulsive disorder. In: Jenike MA, Asberg M, editors. Understanding obsessive-compulsive disorder (OCD). Kirkland, WA: Hogrefe and Huber; 1991:17-23.
7. McKay D, Abramowitz JS, Calamari JE, Kyrios M, Radomsky A, Sookman D, et al. A critical evaluation of obsessive-compulsive disorder subtypes: symptoms versus mechanisms. Clin Psychol Rev 2004 Jul;24(3):283-313. [doi: 10.1016/j.cpr.2004.04.003] [Medline: 15245833]

8. Bloch MH, Landeros-Weisenberger A, Rosario MC, Pittenger C, Leckman JF. Meta-analysis of the symptom structure of obsessive-compulsive disorder. Am J Psychiatry 2008 Dec;165(12):1532-1542 [FREE Full text] [doi: 10.1176/appi.ajp.2008.08020320] [Medline: 18923068]
9. Leckman JF, Grice DE, Boardman J, Zhang H, Vitale A, Bondi C, et al. Symptoms of obsessive-compulsive disorder. Am J Psychiatry 1997 Jul;154(7):911-917. [doi: 10.1176/ajp.154.7.911] [Medline: 9210740]
10. Summerfeldt LJ, Richter MA, Antony MM, Swinson RP. Symptom structure in obsessive-compulsive disorder: a confirmatory factor-analytic study. Behav Res Ther 1999 Apr;37(4):297-311. [doi: 10.1016/s0005-7967(98)00134-x] [Medline: 10204276]
11. Baer L. Factor analysis of symptom subtypes of obsessive compulsive disorder and their relation to personality and tic disorders. J Clin Psychiatry 1994 Mar;55 Suppl:18-23. [Medline: 8077163]
12. Hantouche EG, Lancrenon S. [Modern typology of symptoms and obsessive-compulsive syndromes: results of a large French study of 615 patients]. Encephale 1996 May;22 Spec No 1:9-21. [Medline: 8767023]
13. Mataix-Cols D, Junqué C, Sànchez-Turet M, Vallejo J, Verger K, Barrios M. Neuropsychological functioning in a subclinical obsessive-compulsive sample. Biol Psychiatry 1999 Apr;45(7):898-904. [doi: 10.1016/s0006-3223(98)00260-1]
14. Mataix-Cols D, Rauch SL, Baer L, Eisen JL, Shera DM, Goodman WK, et al. Symptom stability in adult obsessive-compulsive disorder: data from a naturalistic two-year follow-up study. Am J Psychiatry 2002 Feb;159(2):263-268. [doi: 10.1176/appi.ajp.159.2.263] [Medline: 11823269]
15. Calamari JE, Wiegartz PS, Janeck AS. Obsessive–compulsive disorder subgroups: a symptom-based clustering approach. Behav Res Ther 1999 Feb;37(2):113-125. [doi: 10.1016/s0005-7967(98)00135-1]
16. Abramowitz JS, Franklin ME, Schwartz SA, Furr JM. Symptom presentation and outcome of cognitive-behavioral therapy for obsessive-compulsive disorder. J Consult Clin Psychol 2003 Dec;71(6):1049-1057. [doi: 10.1037/0022-006X.71.6.1049] [Medline: 14622080]
17. Calamari JE, Wiegartz PS, Riemann BC, Cohen RJ, Greer A, Jacobi DM, et al. Obsessive–compulsive disorder subtypes: an attempted replication and extension of a symptom-based taxonomy. Behav Res Ther 2004 Jun;42(6):647-670. [doi: 10.1016/s0005-7967(03)00173-6]
18. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. Arch Gen Psychiatry 1989 Nov;46(11):1006-1011. [Medline: 2684084]
19. Doron G, Derby DS, Szepsenwol O. Relationship obsessive compulsive disorder (ROCD): a conceptual framework. J Obsessive Compuls Relat Disord 2014 Apr;3(2):169-180. [doi: 10.1016/j.jocrd.2013.12.005]
20. Doron G, Mizrahi M, Szepsenwol O, Derby D. Right or flawed: relationship obsessions and sexual satisfaction. J Sex Med 2014 Sep;11(9):2218-2224. [doi: 10.1111/jsm.12616] [Medline: 24903281]
21. Doron G, Derby D, Szepsenwol O, Nahaloni E, Moulding R. Relationship obsessive-compulsive disorder: interference, symptoms, and maladaptive beliefs. Front Psychiatry 2016 Apr 18;7:58 [FREE Full text] [doi: 10.3389/fpsyt.2016.00058] [Medline: 27148087]
22. Cerea S, Ghisi M, Bottesi G, Carraro E, Broggio D, Doron G. Reaching reliable change using short, daily, cognitive training exercises delivered on a mobile application: The case of Relationship Obsessive Compulsive Disorder (ROCD) symptoms and cognitions in a subclinical cohort. J Affect Disord 2020 Nov 01;276:775-787. [doi: 10.1016/j.jad.2020.07.043] [Medline: 32738662]
23. Trak E, Inozu M. Developmental and self-related vulnerability factors in relationship-centered obsessive compulsive disorder symptoms: a moderated mediation model. J Obsessive Compuls Relat Disord 2019 Apr;21:121-128. [doi: 10.1016/j.jocrd.2019.03.004]
24. Melli G, Bulli F, Doron G, Carraresi C. Maladaptive beliefs in relationship obsessive compulsive disorder (ROCD): replication and extension in a clinical sample. J Obsessive Compuls Relat Disord 2018 Jul;18:47-53. [doi: 10.1016/j.jocrd.2018.06.005]
25. Mataix-Cols D, Fullana MA, Alonso P, Menchón JM, Vallejo J. Convergent and discriminant validity of the Yale-Brown Obsessive-Compulsive Scale Symptom Checklist. Psychother Psychosom 2004;73(3):190-196. [doi: 10.1159/000076457] [Medline: 15031592]
26. Sulkowski ML, Storch EA, Geffken GR, Ricketts E, Murphy TK, Goodman WK. Concurrent validity of the Yale-Brown Obsessive-Compulsive Scale-Symptom Checklist. J Clin Psychol 2008 Dec;64(12):1338-1351. [doi: 10.1002/jclp.20525] [Medline: 18942133]
27. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019;2:45 [FREE Full text] [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
28. Liddy ED. Natural Language Processing. Encyclopedia of Library and Information Science, 2nd Ed. New York: Marcel Decker, Inc; 2001. URL: https://surface.syr.edu/istpub/63/ [accessed 2021-04-29]
29. NOCD therapy. URL: https://www.treatmyocd.com/ [accessed 2021-05-11]
30. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); October, 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]

31. Dingwall N, Potts C. Mittens: an extension of GloVe for learning domain-specialized representations internet. arXiv - Computer Science; Computation and Language. 2018. URL: http://arxiv.org/abs/1803.09901 [accessed 2021-04-29]
32. Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D. The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2014 Presented at: 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations; June, 2014; Baltimore, Maryland p. 55-60. [doi: 10.3115/v1/p14-5010]
33. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987 Nov;20:53-65. [doi: 10.1016/0377-0427(87)90125-7]
34. Thorndike RL. Who belongs in the family? Psychometrika 1953 Dec;18(4):267-276. [doi: 10.1007/BF02289263]
35. Calinski T, Harabasz J. A dendrite method for cluster analysis. Commun Stat 1974;3(1):1-27. [doi: 10.1080/03610927408827101]
36. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979 Feb;1(2):224-227. [Medline: 21868852]
37. Feinstein SB, Fallon BA, Petkova E, Liebowitz MR. Item-by-item factor analysis of the Yale-Brown Obsessive Compulsive Scale Symptom Checklist. J Neuropsychiatry Clin Neurosci 2003;15(2):187-193. [doi: 10.1176/jnp.15.2.187] [Medline: 12724460]
38. Gandhi M, Wang T. Digital health consumer adoption: 2015. Rock Health. 2021. URL: https://rockhealth.com/reports/digital-health-consumer-adoption-2015/ [accessed 2021-05-11]
39. Rasmussen SA, Eisen JL. The epidemiology and clinical features of obsessive compulsive disorder. Psychiatr Clin North Am 1992 Dec;15(4):743-758. [Medline: 1461792]
40. Ball SG, Baer L, Otto MW. Symptom subtypes of obsessive-compulsive disorder in behavioral treatment studies: a quantitative review. Behav Res Ther 1996 Jan;34(1):47-51. [doi: 10.1016/0005-7967(95)00047-2] [Medline: 8561764]
41. Glazier K, Swing M, McGinn LK. Half of obsessive-compulsive disorder cases misdiagnosed: vignette-based survey of primary care physicians. J Clin Psychiatry 2015 Jun;76(6):761-767. [doi: 10.4088/JCP.14m09110] [Medline: 26132683]
42. Glazier K, Calixte RM, Rothschild R, Pinto A. High rates of OCD symptom misidentification by mental health professionals. Ann Clin Psychiatry 2013 Aug;25(3):201-209. [Medline: 23926575]
43. Coles ME, Frost RO, Heimberg RG, Rhéaume J. “Not just right experiences”: perfectionism, obsessive–compulsive features and general psychopathology. Behav Res Ther 2003 Jun;41(6):681-700. [doi: 10.1016/s0005-7967(02)00044-x]
44. Research Domain Criteria (RDoC). NIMH (Intramural Research Program). URL: https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/ [accessed 2021-05-13]

Abbreviations

GloVe: Global Vectors for Word Representation
NLP: natural language processing
OCD: obsessive-compulsive disorder
YBOCS-SC: Yale-Brown Obsessive-Compulsive Scale Symptom Checklist

Edited by R Kukafka; submitted 03.11.20; peer-reviewed by G Doron, D Mendes; comments to author 29.12.20; revised version received 09.01.21; accepted 19.04.21; published 21.06.21.

Please cite as:
Feusner JD, Mohideen R, Smith S, Patanam I, Vaitla A, Lam C, Massi M, Leow A
Semantic Linkages of Obsessions From an International Obsessive-Compulsive Disorder Mobile App Data Set: Big Data Analytics Study
J Med Internet Res 2021;23(6):e25482
URL: https://www.jmir.org/2021/6/e25482
doi:10.2196/25482
PMID:33892466

Original Paper

Perceptions of Racially and Ethnically Diverse Women at High Risk of Breast Cancer Regarding the Use of a Web-Based Decision Aid for Chemoprevention: Qualitative Study Nested Within a Randomized Controlled Trial

Tarsha Jones[1], MSN, PhD; Ashlee Guzman[2], BS; Thomas Silverman[2], MPH; Katherine Freeman[1], DPH; Rita Kukafka[2], DPH, MA; Katherine Crew[2], MD, MS

[1]Florida Atlantic University, Boca Raton, FL, United States
[2]Columbia University, New York, NY, United States

Corresponding Author:
Tarsha Jones, MSN, PhD
Florida Atlantic University
777 Glades Road
Boca Raton, FL, 33431
United States
Phone: 1 561 297 4975
Email: Jonest@health.fau.edu

Abstract

Background: Chemopreventive agents such as selective estrogen receptor modulators and aromatase inhibitors have proven efficacy in reducing breast cancer risk by 41% to 79% in high-risk women. Women at high risk of developing breast cancer face the complex decision of whether to take selective estrogen receptor modulators or aromatase inhibitors for breast cancer chemoprevention. *RealRisks* is a patient-centered, web-based decision aid (DA) designed to promote the understanding of breast cancer risk and to engage diverse women in planning a preference-sensitive course of decision making about taking chemoprevention.

Objective: This study aims to understand the perceptions of women at high risk of developing breast cancer regarding their experience with using *RealRisks*—a DA designed to promote the uptake of breast cancer chemoprevention—and to understand their information needs.

Methods: We completed enrollment to a randomized controlled trial among 300 racially and ethnically diverse women at high risk of breast cancer who were assigned to standard educational materials alone or such materials in combination with *RealRisks*. We conducted semistructured interviews with a subset of 21 high-risk women enrolled in the intervention arm of the randomized controlled trial who initially accessed the tool (on average, 1 year earlier) to understand how they interacted with the tool. All interviews were audio recorded, transcribed verbatim, and compared with digital audio recordings to ensure the accuracy of the content. We used content analysis to generate themes.

Results: The mean age of the 21 participants was 58.5 (SD 10.1) years. The participants were 5% (1/21) Asian, 24% (5/21) Black or African American, and 71% (15/21) White; 10% (2/21) of participants were Hispanic or Latina. All participants reported using *RealRisks* after being granted access to the DA. In total, 4 overarching themes emerged from the qualitative analyses: the acceptability of the intervention, specifically endorsed elements of the DA, recommendations for improvements, and information needs. All women found *RealRisks* to be acceptable and considered it to be helpful (21/21, 100%). Most women (13/21, 62%) reported that *RealRisks* was easy to navigate, user-friendly, and easily accessible on the web. The majority of women (18/21, 86%) felt that *RealRisks* improved their knowledge about breast cancer risk and chemoprevention options and that *RealRisks* informed their (17/21, 81%) decision about whether or not to take chemoprevention. Some women (9/21, 43%) shared recommendations for improvements, as they wanted more tailoring based on user characteristics, felt that the DA was targeting a narrow population of Hispanic or Latina by using graphic novel–style narratives, wanted more understandable terminology, and felt that the tool placed a strong emphasis on chemoprevention drugs.

Conclusions: This qualitative study demonstrated the acceptability of the *RealRisks* web-based DA among a diverse group of high-risk women, who provided some recommendations for improvement.

(J Med Internet Res 2021;23(6):e23839) doi:10.2196/23839

KEYWORDS

breast cancer; chemoprevention; qualitative; decision support; cancer; estrogen receptor; web-based; cancer risk

Introduction

Breast cancer is the most commonly diagnosed cancer among women in the United States and the second leading cause of cancer-related deaths; therefore, prevention of the disease would significantly improve public health [1]. An estimated 268,600 new cases of invasive breast cancer occur among women each year, and 41,769 women will die from the disease [1,2]. The United States Preventive Services Task Force recommends that clinicians offer chemoprevention as a primary prevention strategy to women at high risk of breast cancer and low risk of adverse effects from these drugs [3]. High-risk criteria for breast cancer are defined as a 5-year invasive breast cancer risk of ≥1.67% or lifetime risk score of ≥20% according to the Gail risk model, which accounts for age, race and ethnicity, benign breast disease, first-degree family history of breast cancer, and reproductive factors [4,5]. Chemopreventive agents such as selective estrogen receptor modulators (tamoxifen and raloxifene) and aromatase inhibitors (exemestane and anastrozole) have been shown to reduce breast cancer risk by 41%-79% among high-risk women [6]. Despite the risk-reducing benefits, chemoprevention uptake remains low in the United States, with fewer than 5% of high-risk women deciding to take the medication [4]. Racial and ethnic minority women are less likely to seek breast cancer preventive care [7,8], which contributes to higher rates of late-stage diagnoses, poorer clinical outcomes, and health disparities [9-11].

Patient-level barriers to chemoprevention uptake include inadequate knowledge and negative attitudes, the fear of potential side effects, and inaccurate perceptions of breast cancer risk [4,12]. Previous studies have found that some women are not aware of the availability of chemoprevention drugs, and less awareness has been noted among racial and ethnic minority women [12]. Women may also be skeptical about the efficacy of chemoprevention in reducing the risk of breast cancer [12]. Studies have found that the fear of the potential side effects of tamoxifen leads to negative attitudes toward chemoprevention, including perceptions that the increased risks of endometrial cancer, pulmonary embolism, stroke, deep vein thrombosis, cataracts, hormonal symptoms, and sexual problems outweigh the potential benefits of the drugs (ie, reduced risks of breast cancer and osteoporosis) [13-16]. In addition, when women view themselves as healthy and not at high risk of developing breast cancer, they are less likely to take chemoprevention drugs [17].

Women at high risk of developing breast cancer face the complex decision of whether to take selective estrogen receptor modulators or aromatase inhibitors for breast cancer chemoprevention [5,12]. The decision is complex because (1) the efficacy of these drugs in preventing breast cancer is limited to estrogen receptor–positive tumors, (2) there is an increased risk of developing serious medical conditions with the use of these drugs, and (3) the recommendations are different for pre- and postmenopausal women [5,12]. Therefore, the best choice for chemoprevention is not always clear, making this a preference-sensitive decision that takes into account how each individual values the relative potential benefits and harms [5,12]. Several interventions have been designed to increase chemoprevention uptake [14,15,18-20]. However, these interventions have been met with limited success, ranging from 0.5% to 5.6% chemoprevention uptake [14,15,18-20]. To address patient-related barriers to chemoprevention among racially and ethnically diverse women, our research team developed a patient-centered, web-based decision aid (DA), *RealRisks*, which is currently being tested in 3 clinical trials [4,21,22].

The purpose of this study is to understand the perceptions of women at high risk of developing breast cancer regarding their experience with using *RealRisks*—a DA designed to promote the uptake of breast cancer chemoprevention—and to understand their information needs.

Methods

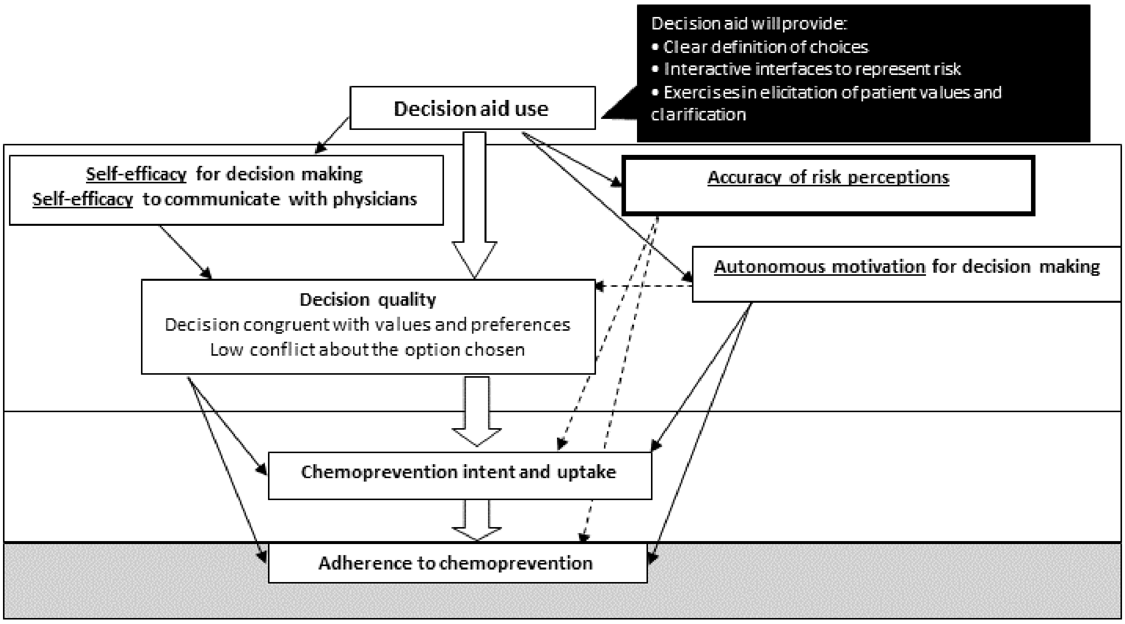
Study Design

We conducted a qualitative study to understand high-risk women's perceptions about *RealRisks*. This study was nested within a randomized controlled trial (RCT) that involved 300 high-risk women assigned to standard educational materials alone or in combination with *RealRisks* to determine chemoprevention uptake at 6 months (primary outcome) [4].

Intervention

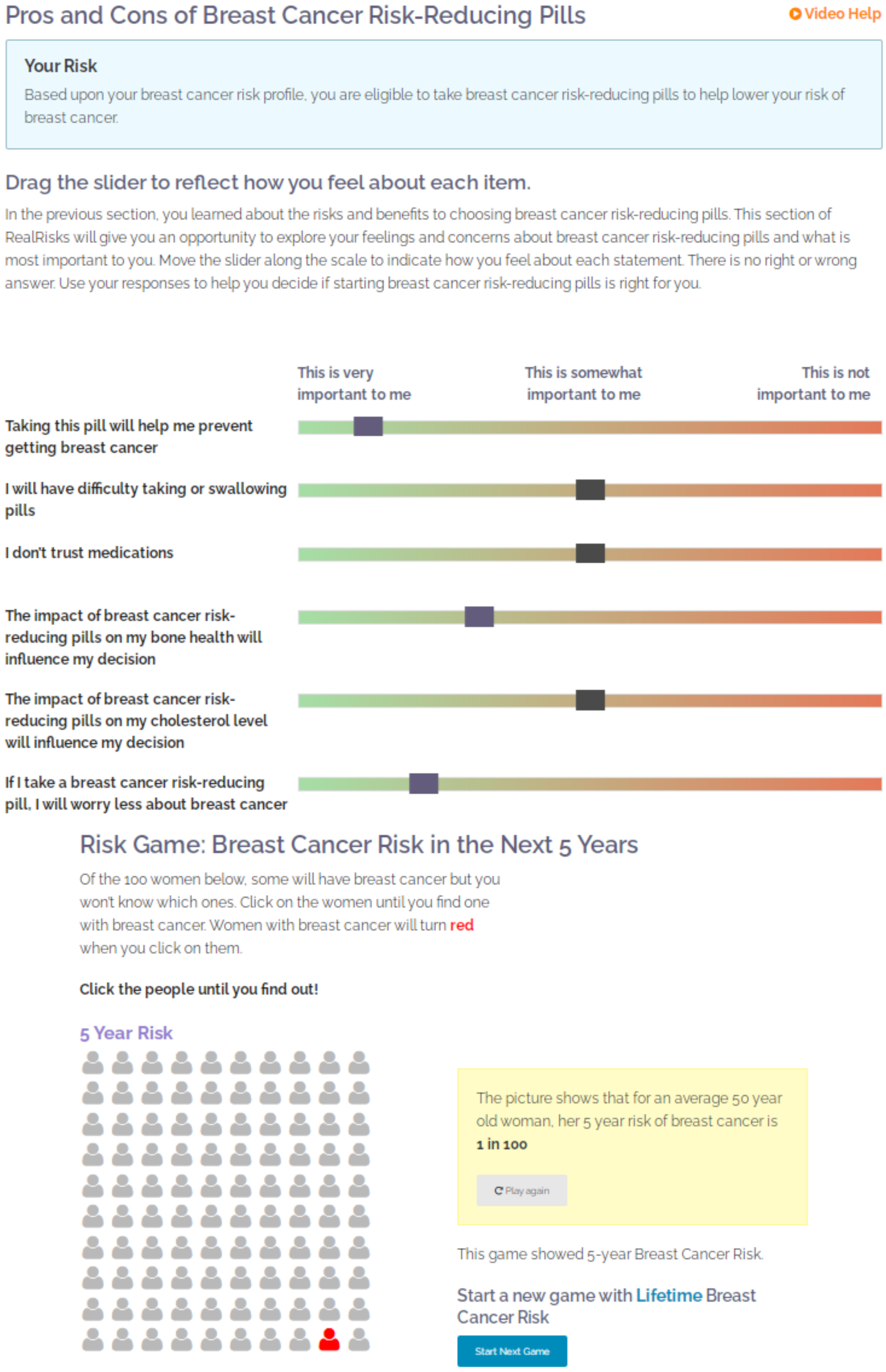
RealRisks was designed to promote a woman's understanding of her breast cancer risk and to engage women in planning a preference-sensitive course of decision making about chemoprevention. This web-based DA incorporates 2 complementary theoretical frameworks—shared decision making [23] and self-determination theory (SDT) [24]—to engage women in planning a preference-sensitive course of decision making about chemoprevention. SDT has at its core the concept of autonomous motivation [24-26] and describes autonomous choices as those that a person could fully endorse upon reflection [27-29]. This is contrasted with behaviors or choices that feel *controlled* or coerced by another person. Fully autonomous choices involve reflecting on and integrating one's preferences and values. The key is to facilitate choice in the context of decisions concerning chemoprevention, in a manner such that the evidence presented is not experienced as coercive but as supportive of autonomous choice. As shown in Figure 1, the *RealRisks* DA is intended to promote the accuracy of breast cancer risk perceptions, autonomous motivation, self-efficacy for decision making, and chemoprevention uptake.

Figure 1. Multidisciplinary framework based on shared decision making and self determination theory.



The DA is delivered via audio files with Spanish translation and is organized into the following modules: (1) breast cancer risk (breast cancer risk factors, calculation of personal breast cancer risk, and interactive games on risk communication), (2) chemoprevention, (3) family history and genetic testing, and (4) lifestyle factors. The content in *RealRisks* has been tailored to a Hispanic or Latina group of women. Figure 2 shows the interactive components of the tool that the women complete during their engagement with the tool. The DA has been described in detail in previous publications [4,22].

Figure 2. Screenshots of RealRisks.



Pros and Cons of Breast Cancer Risk-Reducing Pills

Your Risk

Based upon your breast cancer risk profile, you are eligible to take breast cancer risk-reducing pills to help lower your risk of breast cancer.

Drag the slider to reflect how you feel about each item.

In the previous section, you learned about the risks and benefits to choosing breast cancer risk-reducing pills. This section of RealRisks will give you an opportunity to explore your feelings and concerns about breast cancer risk-reducing pills and what is most important to you. Move the slider along the scale to indicate how you feel about each statement. There is no right or wrong answer. Use your responses to help you decide if starting breast cancer risk-reducing pills is right for you.

This is very important to me | This is somewhat important to me | This is not important to me

Taking this pill will help me prevent getting breast cancer

I will have difficulty taking or swallowing pills

I don't trust medications

The impact of breast cancer risk-reducing pills on my bone health will influence my decision

The impact of breast cancer risk-reducing pills on my cholesterol level will influence my decision

If I take a breast cancer risk-reducing pill, I will worry less about breast cancer

Risk Game: Breast Cancer Risk in the Next 5 Years

Of the 100 women below, some will have breast cancer but you won't know which ones. Click on the women until you find one with breast cancer. Women with breast cancer will turn red when you click on them.

Click the people until you find out!

5 Year Risk

The picture shows that for an average 50 year old woman, her 5 year risk of breast cancer is **1 in 100**

This game showed 5-year Breast Cancer Risk.

Start a new game with Lifetime Breast Cancer Risk

Participants and Settings

Interview participants were recruited from the intervention arm of the RCT. Participants were eligible if they (1) were English or Spanish speaking, (2) completed their 6-month survey, and (3) had access to the web-based *RealRisks* DA through the intervention arm of the RCT. Recruitment and data collection occurred between February and April 2020. Eligible participants were recruited via email and telephone interviews. The research team stopped recruitment efforts upon reaching data saturation—the point at which the collected data stopped producing novel insights [30]. The study protocol was approved by the institutional review board at Columbia University and Florida Atlantic University.

Data Collection

Eligible participants were invited to participate in a one-on-one semistructured interview conducted over the telephone or via

Zoom web conferencing technology (Zoom Video Communications). The interviews were audio recorded to ensure accuracy. A semistructured interview guide was developed with open-ended questions to explore women's perceptions and experiences using *RealRisks*. We explored participants' acceptability of the tool (what they liked or disliked about *RealRisks* and what new information they learned) and any unmet information needs after interacting with the tool. We also explored participants' access to electronic devices used to access the tool and any technological issues that participants may have encountered. In addition, we asked women about their decision making regarding chemoprevention, factors impacting their decisions, and their clinical encounters with their health care provider. A total of 2 versions of the interview guides were developed: one for women who used *RealRisks* and another for women who had never accessed the DA and could not recall using *RealRisks*. The interviews were conducted by TJ and AG and lasted between 20 and 60 minutes. The interviews were transcribed verbatim by a Health Insurance Portability and Accountability Act–compliant transcription company and compared with the original digital recording to ensure the accuracy of the content. This paper reports data on the perceptions, experiences, and information needs of high-risk women who were granted access to the DA through the intervention arm of the parent trial.

Data Analysis

Transcribed and deidentified data were analyzed using Dedoose software (SocioCultural Research Consultants). Content analysis was used to systematically describe the meaning of the qualitative data [31,32]. Data analyses were performed in several steps. First, the principal investigator (TJ) read the first 2 transcripts to gain familiarity with the data and used open coding to build a coding framework derived from the interview guide topics. Second, the analysis team, consisting of researchers trained in conducting qualitative research (TJ, AG, and TS), coded the first 2 transcripts independently using line-by-line coding and discussed code applications as a group to develop consensus. Third, the analysis team members used the Dedoose training center to evaluate interrater reliability by generating a pooled Cohen κ coefficient to assess coding precision [33]. The final pooled Cohen κ score was 0.86, which indicated a high level of coding agreement among the coding team [34]. Fourth, TJ and AG continued to code the remaining transcripts independently and used the memo feature in Dedoose to review each other's coding and to capture thoughts and observations within and between transcripts. Finally, the analysis team abstracted and interpreted the data to generate thematic domains. To ensure trustworthiness, direct quotations were provided to connect the results to the raw data. Ellipses were used (3 periods indicating a break within a sentence) to help minimize the length of the quotations. In addition, some larger quotations were presented to keep the context of the conversation intact.

Results

Participants and Interview Data

A total of 21 high-risk women participated in this qualitative study (Table 1). The mean age of the participants was 58.5 (SD 10.1) years. Our sample was racially and ethnically diverse, with demographics distributed as follows: 5% (1/21) Asian, 24% (5/21) Black or African American, and 71% (15/21) White. In addition, 2 participants were Hispanic or Latina. Most (15/21, 71%) participants had a family history of breast cancer, and all women (21/21, 100%) reported using *RealRisks* after being granted access to the DA. Interview data were categorized into 4 main themes: (1) the acceptability of the intervention, (2) specifically endorsed elements of the DA, (3) recommendations for improvement, and (4) information needs. Each theme comprised several subthemes, and exemplar quotes are presented to authenticate the overarching themes and subthemes. Themes, subthemes, and exemplar quotes are also provided in Multimedia Appendix 1.

Table 1. Study participant characteristics (N=21).

Characteristics	Value
Age (years), mean (SD)	58.5 (10.1)
Ethnicity, n (%)	
Hispanic or Latina	2 (9)
Non-Hispanic	19 (91)
Race, n (%)	
Asian	1 (5)
Black or African American	5 (24)
White	15 (71)
Highest level of education, n (%)	
High school or GED[a]	2 (11)
Associates or bachelors	9 (50)
Some college	2 (11)
Graduate degree	5 (28)
Primary language, n (%)	
English	21 (100)
Family history of breast cancer, n (%)	
Yes	15 (71)
No	6 (29)
Taking chemoprevention, n (%)	
Yes	20 (95)
No	1 (5)
Used the RealRisks decision aid, n (%)	
Yes	21 (100)
No	0 (0)

[a] GED: General Education Development.

Theme 1: The Acceptability of the Intervention

Subtheme 1: General Perceptions

All women (21/21, 100%) who accessed *RealRisks* shared their general perceptions that they liked using the DA and considered it to be helpful. Almost 80% (16/21, 76%) of the women who were granted access to the DA through the intervention arm of the parent trial completed the tool. Some women (9/21, 43%) shared recommendations for improvement after using the tool. Overall, women from all 4 racial and ethnic groups represented in our study viewed the tool as acceptable. One of the women stated:

> *I felt that it was good for me to have done it [RealRisks]. And I do feel that I got information that I would not have gotten otherwise. I mean, I felt a little bit more informed. So, I'm happy about that.* [Participant #18]

Subtheme 2: Usability of the Intervention

Most women (13/21, 62%) felt that *RealRisks* was easy to navigate. For instance, women shared the following:

> *It was pretty simple and easy to do, and I didn't feel like there were a lot of issues with using the tool and taking the surveys at all...I didn't find it to be off putting. I didn't find it to be hard to do. I didn't find it to suck up a lot of my time.* [Participant #27]

> *It was easy to access. The instructions were pretty good...Simple and basic and not overwhelming and I like the fact that if you didn't know something you could just hover over a term and you could get a definition.* [Participant #26]

> *I liked the fact that it was quick. I was able to just go through it.* [Participant #21]

> *I thought it was easy to use. It answered the questions that I would have. Like it was very thorough in the questions, so I didn't have to wonder what it meant. It wasn't up for interpretation. I thought it was pretty clear.* [Participant #25]

None of the women who used the DA reported encountering any technological issues. One woman stated:

> *No. I was able to get in. I was able to answer all the questions. I don't remember any glitches at all.* [Participant #15]

Subtheme 3: User-friendly

Other women felt that the DA was user-friendly, straightforward, simple to follow, and easy to understand. They shared the following:

> *It was pretty easy to understand. I have no medical background and it gave, I thought, very clear, information about how different drugs could help. And then I could answer the questions that came next.* [Participant #23]
>
> *Pretty easy. Sort of user-friendly in terms of just accessing all the steps.* [Participant #18]
>
> *It was simple to follow-simple to use. It didn't take up too much time. I mean, it's been a while. It's been over a year, right? But I remember just being able to just quickly go on my tablet and answer the questions.* [Participant #7]
>
> *I thought it was easy to understand and to complete. I thought it was pretty user-friendly for someone like myself.* [Participant #5]

Subtheme 4: Easily Accessible on the Web

Most women (19/21, 90%) reported having access to technological devices such as cell phones, tablets, laptops, and desktop computers to access *RealRisks.* The exceptions were 2 Spanish-speaking participants who reported having no access to a computer at home and came to the hospital to receive in-person support to access the tool from one of our research team members. The women who had access to technology liked that the DA was web-based, which allowed them to access it on the web at their own convenience. For example, women shared the following:

> *It was easy because I could do it online. You know, I could like start it at work. I could finish it at home. I did not have to go back and re-answer questions. I thought it was user-friendly.* [Participant #1]
>
> *I could do it online, could start it at work and finish it at home.* [Participant #2]
>
> *It was good that it was online and it was easy to navigate.* [Participant #14]
>
> *I didn't really have any trouble following through. But, you know, I guess, with everything that's done online, you're doing it on your own time.* [Participant #18]

Theme 2: Specifically Endorsed Elements of the DA

Subtheme 1: Improved Knowledge

Most women (18/21, 86%) described gaining new knowledge after using *RealRisks*. A total of 11 women described learning new information about breast cancer prevention drugs. The participants described learning new information such as which chemoprevention drug was appropriate for them, for example, tamoxifen and other available drugs. Other women described learning which drug was appropriate for premenopausal and postmenopausal women. The women described the following:

> *New information was the information around the drugs to take. I can't remember what they were called, but the preventive, I guess that's what they are, drugs.* [Participant #5]
>
> *I knew about the existence of the drug, tamoxifen. I know it opposes estrogen. An anti-estrogen-type drug. I did not know about other drugs. So that was new information.* [Participant #24]

In total, 2 of these women explained that *RealRisks* reinforced some of the information that they had already heard from their health care providers, which contributed to an improvement in their overall knowledge:

> *Being diagnosed with the BRCA2 gene mutation, I was getting so much information at that time and doing so much research and meeting with so many different doctors because of trying to be very, very proactive about being diagnosed with the BRCA2 gene that it actually turned out to have been helpful in terms of reinforcing some ideas that I didn't necessarily—that hadn't quite sunk in yet.* [Participant #8]

A total of 8 women described learning about their *risk* of breast cancer. These women stated that they now understood the factors contributing to *risk*, such as genetics, personal history, family history, and lifestyle factors. One woman appreciated the way *RealRisks* presented *risk* information in a story format. Another woman who had a pathogenic variant in the *BRCA*2 gene felt that *RealRisks* was thought provoking. The women learned actions that they could take to reduce their risk and shared the following:

> *Oh, it made me think about what the risks are, specifically having to do with my situation and being in a high-risk category...Well, I guess I surmised that I was high risk, given my family history. But it solidified that, and it also taught me about options that were available to me should I choose them to reduce my risk.* [Participant #15]
>
> *I like the way that it was set up, as a possibility of being at risk, and that was for me, a sensitive issue...I remember there was a story—the way that it was presented by story. So it became more real rather than just the data, medical data, and to try to identify through the just numbers. It was like real.* [Participant #17]

Subtheme 2: Informed Decision Making

Most women (17/21, 81%) who used *RealRisks* reported that the DA played a role in informing their decision making about chemoprevention. The women reviewed the modules in *RealRisks* and became familiar with the options that exist to reduce their breast cancer risk, including developing an understanding of the benefits and risks of taking a chemoprevention pill. The women explained:

> *Yes, because the information that was provided was thorough. I mean, there was plenty of it. And I feel comfortable that—I've seen both sides of it. I could see why a person would take it. I believe I have made an informed decision.* [Participant #6]
>
> *I think it is because I did not know about it [chemoprevention] until I saw RealRisks. But definitely it played into it [my decision-making]. And I didn't really look into it because I thought I wasn't really the best person to be eligible for it. And now I know I do have an elevated risk.* [Participant #20]

Although 95% (20/21) of the women in the study viewed themselves as *healthy* and decided not to take chemoprevention, their experience with using *RealRisks* informed an autonomous decision about whether chemoprevention was the right choice for them:

> *I made a decision for who I am and what I want right now. So I didn't ask my doctor. I didn't ask anyone. I just made the decision by myself. I just took my decision because I don't want to put anything in my body. That's it.* [Participant #9]
>
> *I remember when they talked about the differing types of procedures you could go through to help prevent breast cancer. It was in-depth and I read it all but I was not in agreement with taking any kind of medicine...It was very informative, and I think if I was in poorer health, I would probably be willing to try anything. But my health is relatively good.* [Participant #22]

Theme 3: Recommendations for Improvement

In addition to positive perceptions about *RealRisks*, some women (9/21, 43%) pointed out aspects of the DA that they did not like and shared recommendations for improvement.

Subtheme 1: Tailored for a Hispanic or Latina Population

Some women (4/9, 44%) who were not of Hispanic or Latina descent were concerned that *RealRisks* appeared to be tailored for Hispanic or Latina women. The women recognized that it was difficult to design a tool for diverse target audiences, especially those with varying health literacy and numeracy. One woman felt that *RealRisks* was designed for people who were not native English speakers. Another woman who was not of Hispanic descent was confused as to why Hispanic characters were used in the illustrations. In addition, 3 women shared that the comics were not appealing to them:

> *It was trying so hard to be ethnically diverse. It felt like it might have been for people where English was not their original language.* [Participant #8]
>
> *Found it interesting that if you're going to draw pictures of—I actually discussed this with my doctor. If you're going to draw pictures of people, it's hard not to racially profile. Even though it was a line drawing, it clearly looked like a person of color. So I mean, not that it meant any good or bad to me.* [Participant #24]

A few women (3/9, 33%) felt that the DA made assumptions about where they were in the care continuum and did not give them the opportunity to input their own risk factors. For instance, women shared the following:

> *There was not an opportunity to discuss why I didn't think I was at risk. I felt that there were assumptions made about me without even simply asking. For example, cigarette smoking causes a great deal of cancers and I do not smoke.* [Participant #6]
>
> *Some of it did not apply to me because I was already under care. I was already diagnosed with a BRCA2 gene mutation and already had surgery and already in the middle of this. But it wasn't quite flexible enough...I sort of stepped into the RealRisks halfway through if you know—I didn't need to assess my risk.* [Participant #8]

Subtheme 2: Difficult Terminology

In total, 33% (3/9) of women felt that *RealRisks* included difficult terminology, and the content was not written in layman's terms. Of these 3 women, 2 (66%) stated that if they were not intelligent, it would have been difficult to understand the information in *RealRisks*. The participants explained the following:

> *I did find some of the terminology a little difficult, because I wasn't familiar with some things. I do remember like, having to look some things up, particularly when it came to the treatment options.* [Participant #7]
>
> *The language used in RealRisks was not layman's language. If I wasn't intelligent, I wouldn't have understood it all. Maybe it was like reading The Times paper.* [Participant #22]
>
> *For me, being a white, middle-class, Jewish college-educated woman, I had no trouble with it. But I wouldn't want to speak for other cultures. I am married to a Puerto Rican man and like I could see where his mother, if she were reading this, would've really—she might've struggled...I think it tried to be very clear and very easy to understand, but I don't know, I think it can be hard for people—sometimes the way people have a block on math, sometimes when they read about medical things they get a bit of a block.* [Participant #23]

Subtheme 3: Strong Emphasis on Chemoprevention

Overall, 22% (2/9) of women felt that the *RealRisks* DA placed a strong emphasis on chemoprevention drugs. Despite the importance of these drugs, these 2 women felt that they would have preferred more emphasis on the lifestyle and environmental factors affecting breast cancer risk:

> *I have two graduate degrees. I'm not saying this to brag. I taught college for many years. So I thought the graphics and the cartoons and all of that, were useful in general, not necessarily for myself. I felt a little bit that the chemoprevention was being pushed.* [Participant #5]

> *I felt, throughout the whole process that it was very much pushing some type of drug on me, that I felt I wasn't really ready to consider that.* [Participant #24]

Theme 4: Information Needs

Participants were asked "what additional information do you need to make decisions about reducing your breast cancer risk?" With respect to women's information needs, 33% (7/21) of women described at least one unmet need. One woman wanted to learn more about mammography screening:

> *I was having this conversation with my sister recently. Because so much of the information around mammography and the frequency of mammography is changing. And so I was just having this conversation with her, that, you know, should I still be getting a yearly mammogram, I know that, age-wise, I think someone my age, and probably, with our risk, we should be getting it. But I feel like, because a lot of the information is changing, there is some confusion in terms of trying to make a decision about certain things.* [Participant #5]

Other women were interested in learning more about breast cancer risk factors, including modifiable lifestyle factors. Some women still desired to understand their individual breast cancer risk and what puts a woman at high risk of developing breast cancer. For example, women shared the following:

> *Well, I think the only information I would need or want would be during the course of my visits with my doctor is to go over my lifestyle, and just to find out if there's any additional information that's become available. And if there's any additional thoughts on the benefits or risks of taking the medications. I would always consider it as I get older.* [Participant #26]
>
> *It's a good question. I guess, printed materials probably, maybe like how to avoid putting myself in harm's way, stuff that you can prevent. You know, I mean, of course there's a lot of stuff that you can't prevent. But I mean, if it proves that, if there was a reason for them to say, "Oh, yeah, you're at really high risk," I, you know, I would do whatever and apply what was suggested.* [Participant #4]
>
> *Probably just you know, a one-page outline of all the risks, you know, maybe just outlining everything in a concise way so that I could look at it and just know what the risks are and just have it—just be more conscious of what the risks are on a day-to-day basis and without having to go through a lot of reading and stuff.* [Participant #18]

Discussion

Principal Findings

In our racially and ethnically diverse sample of women at high risk of developing breast cancer, all participants reported that they used the web-based DA *RealRisks*. The high-risk women who participated in the intervention arm of the chemoprevention trial and were granted access to the DA reported that they generally liked the DA and considered it to be helpful. This result confirms the need for decision support for women at a high risk of developing breast cancer. In line with our objective to understand the perceptions of women at high risk of developing breast cancer regarding their experience with using *RealRisks*, our analysis generated 4 themes:

1. The acceptability of the intervention: women found the tool to be acceptable, easy to navigate, user-friendly, and easily accessible on the web.
2. Specifically endorsed elements of the DA: the DA was favored for its ability to improve knowledge and inform decision making.
3. Recommendations for improvement: participants wanted more tailoring based on user characteristics.
4. Information needs: participants reported wanting to learn more about mammography screening and breast cancer risk factors, including modifiable lifestyle factors.

Regarding theme 1, participants found *RealRisks* to be an acceptable intervention and had positive attitudes overall. The web-based nature of the intervention was very appealing to participants who appreciated being able to access the tool at their own convenience. In addition, the tool was acceptable to women in a variety of age groups, as the average age of our participants was 58.6 years (SD 10.1). Regarding theme 2, the women reported that *RealRisks* increased their knowledge about chemoprevention. This finding is aligned with the results of our pilot study, which found an increase in chemoprevention knowledge after exposure to *RealRisks* [22]. The observation that most women were autonomously motivated to make an informed decision to not take chemoprevention after using the *RealRisks* DA is aligned with the SDT, the underlining premise of the DA [35]. SDT, a theory of human motivation, emphasizes the extent to which behaviors are relatively autonomous, originating from oneself without pressure or coercion by interpersonal forces [35]. Therefore, implementing the DA in primary care may not be appropriate for all high-risk women and should be reserved for women who are undecided about taking chemoprevention and would like to further discuss the pros and cons and chemoprevention options with their primary care provider. Our results indicate that the DA offered great decision support to high-risk women, improved their knowledge, and informed their decision making about chemoprevention.

Regarding theme 3, despite the general acceptability of the DA, high-risk women wanted a more holistic approach to reducing their breast cancer risk based on tailoring that incorporates user characteristics [36]. The women felt that the DA was not personalized and made assumptions about where they were in the care continuum, which did not give them the opportunity to input their own risk factors. Other women felt that the DA placed a strong emphasis on chemoprevention drugs. Regarding theme 4, our results demonstrate that, based on the information needs reported, the high-risk women enrolled in our study wanted to learn more about general breast cancer risk factors and breast cancer screening strategies, including mammographic frequency, in addition to chemoprevention. Existing evidence suggests that interventions that are tailored based on unique characteristics and provide personalized feedback, guidance, and motivation to users might assist them in engaging in more active lifestyles [37-39] and are an effective method of promoting mammography adherence [40]. Therefore, based on our findings, future iterations of *RealRisks* may consider incorporating tailoring with a broader focus on breast cancer risk reduction and should incorporate guidance on other breast cancer risk factors.

Strengths and Limitations

Our study has several important strengths and limitations. A strength of this study is that we had about 40% (8/21, 38%) racial and ethnic minorities, which is representative of the US population. However, we only included participants from the intervention arm of the chemoprevention trial who had accessed the tool, on average, 1 year earlier and, therefore, may have experienced recall bias. In addition, this study was conducted at a large urban academic medical center, which may not be generalizable to other geographic and practice settings.

Future Work

Future studies should focus on developing interventions that target the knowledge gaps identified in this study and meet the needs of high-risk women to empower them to make informed decisions about reducing their breast cancer risk. Our research group is currently developing a breast cancer screening module to be added to *RealRisks*, which will provide evidence-based information on mammography screening and other methods of breast imaging.

Conclusions

This study has demonstrated the acceptability of the *RealRisks* web-based DA among a diverse group of high-risk women, with a few caveats and recommendations for improvement. Women's perceptions about *RealRisks* were influenced by the tool's ability to increase their knowledge about breast cancer risk and the pros and cons of taking chemoprevention drugs; they indicated that this information facilitated their informed decision making about taking chemoprevention. We found that after using the DA, women wanted a more tailored and holistic approach to reducing breast cancer risk. The findings will inform the future development of the web-based DA. Next steps include translating the results from this study into further development of the DA and optimizing the tool's module architecture to add a breast cancer screening module.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Themes, subthemes, and exemplar quotes.

[DOCX File , 14 KB - jmir_v23i6e23839_app1.docx]

References

1. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Sauer AG, et al. Breast cancer statistics, 2019. CA Cancer J Clin 2019 Nov 02;69(6):438-451 [FREE Full text] [doi: 10.3322/caac.21583] [Medline: 31577379]
2. American Cancer Society. Breast cancer facts and figures-2020. 2019. URL: https://www.cancer.org/research/cancer-facts-statistics/breast-cancer-facts-figures.html [accessed 2020-05-15]
3. US Preventive Services Task Force, Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, et al. Medication use to reduce risk of breast cancer: US Preventive Services Task Force recommendation statement. J Am Med Assoc 2019 Sep 03;322(9):857-867. [doi: 10.1001/jama.2019.11885] [Medline: 31479144]
4. Crew KD, Silverman TB, Vanegas A, Trivedi MS, Dimond J, Mata J, et al. Study protocol: randomized controlled trial of web-based decision support tools for high-risk women and healthcare providers to increase breast cancer chemoprevention. Contemp Clin Trials Commun 2019 Dec;16:100433 [FREE Full text] [doi: 10.1016/j.conctc.2019.100433] [Medline: 31497674]
5. Ropka ME, Keim J, Philbrick JT. Patient decisions about breast cancer chemoprevention: a systematic review and meta-analysis. J Clin Oncol 2010 Jun 20;28(18):3090-3095. [doi: 10.1200/jco.2009.27.8077]
6. Farkas A, Vanderberg R, Merriam S, DiNardo D. Breast cancer chemoprevention: a practical guide for the primary care provider. J Womens Health (Larchmt) 2020 Jan 01;29(1):46-56. [doi: 10.1089/jwh.2018.7643] [Medline: 31560601]
7. Kaplan CP. Breast cancer risk reduction options: awareness, discussion, and use among women from four ethnic groups. Cancer Epidemiol Biomarkers Prev 2006 Jan 01;15(1):162-166. [doi: 10.1158/1055-9965.epi-04-0758]
8. Jacobson JS, Grann VR, Hershman D, Troxel AB, Li H, Neugut AI. Breast biopsy and race/ethnicity among women without breast cancer. Cancer Detect Prev 2006 Jan;30(2):129-133. [doi: 10.1016/j.cdp.2006.02.002] [Medline: 16621329]
9. Armstrong K, Micco E, Carney A, Stopfer J, Putt M. Racial differences in the use of BRCA1/2 testing among women with a family history of breast or ovarian cancer. J Am Med Assoc 2005 Apr 13;293(14):1729-1736. [doi: 10.1001/jama.293.14.1729] [Medline: 15827311]
10. Hall MJ, Olopade OI. Disparities in genetic testing: thinking outside the BRCA box. J Clin Oncol 2006 May 10;24(14):2197-2203. [doi: 10.1200/jco.2006.05.5889]
11. Ward EM, Smith RA. Integrating tools for breast cancer risk assessment, risk reduction, and early detection. Cancer Epidemiol Biomarkers Prev 2010 Oct 06;19(10):2428-2429. [doi: 10.1158/1055-9965.epi-10-0835]

12. Korfage IJ, Fuhrel-Forbis A, Ubel PA, Zikmund-Fisher BJ, Greene SM, McClure JB, et al. Informed choice about breast cancer prevention: randomized controlled trial of an online decision aid intervention. Breast Cancer Res 2013 Sep 3;15(5):74 [FREE Full text] [doi: 10.1186/bcr3468] [Medline: 24004815]
13. Bober SL, Hoke LA, Duda RB, Regan MM, Tung NM. Decision-making about tamoxifen in women at high risk for breast cancer: clinical and psychological factors. J Clin Oncol 2004 Dec 15;22(24):4951-4957. [doi: 10.1200/jco.2004.05.192]
14. Port ER, Montgomery LL, Heerdt AS, Borgen PI. Patient reluctance toward tamoxifen use for breast cancer primary prevention. Ann Surg Oncol 2001 Aug;8(7):580-585. [doi: 10.1007/s10434-001-0580-9] [Medline: 11508619]
15. Fagerlin A, Dillard AJ, Smith DM, Zikmund-Fisher BJ, Pitsch R, McClure JB, et al. Women's interest in taking tamoxifen and raloxifene for breast cancer prevention: response to a tailored decision aid. Breast Cancer Res Treat 2011 Jun 26;127(3):681-688 [FREE Full text] [doi: 10.1007/s10549-011-1450-1] [Medline: 21442198]
16. Melnikow J, Paterniti D, Azari R, Kuenneth C, Birch S, Kuppermann M, et al. Preferences of Women Evaluating Risks of Tamoxifen (POWER) study of preferences for tamoxifen for breast cancer risk reduction. Cancer 2005 May 15;103(10):1996-2005 [FREE Full text] [doi: 10.1002/cncr.20981] [Medline: 15825209]
17. Salant T. "Why Take It If You Don't Have Anything?" Breast cancer risk perceptions and prevention choices at a public hospital. J Gen Intern Med 2006 Jul;21(7):779-785. [doi: 10.1111/j.1525-1497.2006.00461_1.x]
18. Taylor R, Taguchi K. Tamoxifen for breast cancer chemoprevention: low uptake by high-risk women after evaluation of a breast lump. Ann Fam Med 2005 May 01;3(3):242-247 [FREE Full text] [doi: 10.1370/afm.284] [Medline: 15928228]
19. Fagerlin A, Zikmund-Fisher BJ, Nair V, Derry HA, McClure JB, Greene S, et al. Women's decisions regarding tamoxifen for breast cancer prevention: responses to a tailored decision aid. Breast Cancer Res Treat 2010 Feb 12;119(3):613-620 [FREE Full text] [doi: 10.1007/s10549-009-0618-4] [Medline: 19908143]
20. Loehberg CR, Jud SM, Haeberle L, Heusinger K, Dilbat G, Hein A, et al. Breast cancer risk assessment in a mammography screening program and participation in the IBIS-II chemoprevention trial. Breast Cancer Res Treat 2010 May 21;121(1):101-110. [doi: 10.1007/s10549-010-0845-8] [Medline: 20306293]
21. Silverman TB, Vanegas A, Marte A, Mata J, Sin M, Ramirez JC, et al. Study protocol: a cluster randomized controlled trial of web-based decision support tools for increasing BRCA1/2 genetic counseling referral in primary care. BMC Health Serv Res 2018 Aug 13;18(1):633 [FREE Full text] [doi: 10.1186/s12913-018-3442-x] [Medline: 30103738]
22. Kukafka R, Fang J, Vanegas A, Silverman T, Crew KD. Pilot study of decision support tools on breast cancer chemoprevention for high-risk women and healthcare providers in the primary care setting. BMC Med Inform Decis Mak 2018 Dec 17;18(1):134 [FREE Full text] [doi: 10.1186/s12911-018-0716-5] [Medline: 30558581]
23. Edwards A, Naik G, Ahmed H, Elwyn GJ, Pickles T, Hood K, et al. Personalised risk communication for informed decision making about taking screening tests. Cochrane Database Syst Rev 2013 Feb 28;2(2):CD001865 [FREE Full text] [doi: 10.1002/14651858.CD001865.pub3] [Medline: 23450534]
24. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. Switzerland: Springer; 1985:1-371.
25. Bidassie B. Breast cancer risk assessment and chemoprevention use among veterans affairs primary care providers: a national online survey. Mil Med 2020;185(3-4):512-518. [doi: 10.1093/milmed/usz291]
26. Stevanato KP, Pedroso RB, Iora P, Santos LD, Pelloso FC, Melo WA, et al. Comparative analysis between the Gail, Tyrer-Cuzick and BRCAPRO Models for breast cancer screening in Brazilian population. Asian Pac J Cancer Prev 2019 Nov 01;20(11):3407-3413. [doi: 10.31557/apjcp.2019.20.11.3407]
27. Moller AC, Ryan RM, Deci EL. Self-determination theory and public policy: improving the quality of consumer decisions without using coercion. J Public Policy Mark 2018 Oct 02;25(1):104-116. [doi: 10.1509/jppm.25.1.104]
28. Deci EL, Ryan RM. The "What" and "Why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
29. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]
30. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018 Sep 14;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
31. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov 01;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
32. Flick U. An Introduction to Qualitative Research. Thousand Oaks, California, United States: SAGE Publications Ltd; 2014:1-616.
33. De Vries H, Elliott MN, Kanouse DE, Teleki SS. Using pooled kappa to summarize interrater agreement across many items. Field Methods 2008 Mar 03;20(3):272-282. [doi: 10.1177/1525822x08317166]
34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
35. Patrick H, Williams GC. Self-determination theory: its application to health behavior and complementarity with motivational interviewing. Int J Behav Nutr Phys Act 2012 Mar 02;9(1):18 [FREE Full text] [doi: 10.1186/1479-5868-9-18] [Medline: 22385676]

36. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun 01;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
37. Monteiro-Guerra F, Signorelli GR, Tadas S, Zubiete ED, Romero OR, Fernandez-Luque L, et al. A personalized physical activity coaching app for breast cancer survivors: design process and early prototype testing. JMIR Mhealth Uhealth 2020 Jul 15;8(7):e17552 [FREE Full text] [doi: 10.2196/17552] [Medline: 32673271]
38. Phillips SM, Conroy DE, Keadle SK, Pellegrini CA, Lloyd GR, Penedo FJ, et al. Breast cancer survivors' preferences for technology-supported exercise interventions. Support Care Cancer 2017 Oct 3;25(10):3243-3252 [FREE Full text] [doi: 10.1007/s00520-017-3735-3] [Medline: 28470368]
39. Roberts AL, Potts HW, Koutoukidis DA, Smith L, Fisher A. Breast, prostate, and colorectal cancer survivors' experiences of using publicly available physical activity mobile apps: qualitative study. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e10918 [FREE Full text] [doi: 10.2196/10918] [Medline: 30609982]
40. Sohl SJ, Moyer A. Tailored interventions to promote mammography screening: a meta-analytic review. Prev Med 2007 Oct;45(4):252-261 [FREE Full text] [doi: 10.1016/j.ypmed.2007.06.009] [Medline: 17643481]

Abbreviations

DA: decision aid
RCT: randomized controlled trial
SDT: self-determination theory

Edited by G Eysenbach; submitted 25.08.20; peer-reviewed by F Holland, J Zhu; comments to author 16.09.20; revised version received 12.10.20; accepted 13.04.21; published 08.06.21.

Please cite as:
Jones T, Guzman A, Silverman T, Freeman K, Kukafka R, Crew K
Perceptions of Racially and Ethnically Diverse Women at High Risk of Breast Cancer Regarding the Use of a Web-Based Decision Aid for Chemoprevention: Qualitative Study Nested Within a Randomized Controlled Trial
J Med Internet Res 2021;23(6):e23839
URL: https://www.jmir.org/2021/6/e23839
doi:10.2196/23839
PMID:34100769

Original Paper

Calls to Action (Mobilizing Information) on Cancer in Online News: Content Analysis

Thomas Hongjie Zhang[1], BA; Jen Sern Tham[1], BA, PhD

[1]Department of Communication, Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Serdang, Malaysia

Corresponding Author:
Jen Sern Tham, BA, PhD
Department of Communication
Faculty of Modern Languages and Communication
Universiti Putra Malaysia
Jalan Universiti 1
Serdang
Malaysia
Phone: 60 397698799
Email: jstham@upm.edu.my

Abstract

Background: The health belief model explains that individual intentions and motivation of health behaviors are mostly subject to external cues to action, such as from interpersonal communications and media consumptions. The concept of mobilizing information (MI) refers to a type of mediated information that could call individuals to carry out particular health actions. Different media channels, especially digital media outlets, play an essential role as a health educator to disseminate cancer health information and persuade and mobilize cancer prevention in the community. However, little is known about calls to action (or MI) in online cancer news, especially from Asian media outlets.

Objective: This study aimed at analyzing cancer news articles that contain MI and their news components on the selected Malaysian English and Chinese newspapers with online versions.

Methods: The Star Online and Sin Chew Online were selected for analysis because the two newspaper websites enjoy the highest circulation and readership in the English language and the Chinese language streams, respectively. Two bilingual coders searched the cancer news articles based on sampling keywords and then read and coded each news article accordingly. Five coding variables were conceptualized from previous studies (ie, cancer type, news source, news focus, cancer risk factors, and MI), and a good consistency using Cohen kappa was built between coders. Descriptive analysis was used to examine the frequency and percentage of each coding item; chi-square test (confidence level at 95%) was applied to analyze the differences between two newspaper websites, and the associations between variables and the presence of MI were examined through binary logistic regression.

Results: Among 841 analyzed news articles, 69.6% (585/841) presented MI. News distributions were unbalanced throughout the year in both English and Chinese newspaper websites; some months occupied peaks (ie, February and October), but cancer issues and MI for cancer prevention received minimal attention in other months. The news articles from The Star Online and Sin Chew Online were significantly different in several news components, such as the MI present rates (χ^2=9.25, P=.003), providing different types of MI (interactive MI: χ^2=12.08, P=.001), interviewing different news sources (government agency: χ^2=12.05, P=.001), concerning different news focus (primary cancer prevention: χ^2=10.98, P=.001), and mentioning different cancer risks (lifestyle risks: χ^2=7.43, P=.007). Binary logistic regression results reported that online cancer news articles were more likely to provide MI when interviewing nongovernmental organizations, focusing on topics related to primary cancer prevention, and highlighting lifestyle risks (odds ratio [OR] 2.77, 95% CI 1.89-4.05; OR 97.70, 95% CI 46.97-203.24; OR 186.28; 95% CI 44.83-773.96; P=.001, respectively).

Conclusions: This study provided new understandings regarding MI in cancer news coverage. This could wake and trigger individuals' preexisting attitudes and intentions on cancer prevention. Thus, health professionals, health journalists, and health campaign designers should concentrate on MI when distributing health information to the community.

(J Med Internet Res 2021;23(6):e26019) doi:10.2196/26019

KEYWORDS

mobilizing information; online cancer news; quantitative content analysis; Malaysia; online news; cancer; infodemiology; media; digital media; digital health; health information; cancer health information

Introduction

Cancer is a pressing public health issue and does not discriminate whether in a developed or developing country [1,2]. Cancer, in fact, remains the second leading cause of death throughout the world. The World Health Organization [1] reported that in 2018, 9.6 million people were estimated to have died from cancer. In other words, 1 in 6 deaths is from cancer of some type. Lung, prostate, colorectal, stomach, and liver cancer are the most common types of cancers in men, while breast, colorectal, lung, cervical, and thyroid cancer are the most common among women [1]. According to a recent report from the International Agency for Research on Cancer [3], breast cancer occupied the first leading incidence rate in Malaysia (17.3%), followed by colorectal cancer (14.0%) and lung cancer (10.7%). Local cancer incidence was not equal across ethnic groups. In recent years, Malaysian Chinese have had the highest cancer incidence rate for the majority of leading cancers. It is more elevated than Malays and Indians [4,5].

Scientific findings showed that at least half of all cancers occurring today could be prevented if individuals adopted healthy lifestyles and behaviors, and many other cancers can be diagnosed through medical interventions in the early stages [6]. Hence, it is crucial to promote cancer prevention and early detection to the individuals. Essentially, cancer prevention is no longer treated as a health issue alone but a developmental issue for a nation. Health promotion regarding cancer issues involves collaborations from different stakeholders, which include conventional media and digital media.

The media provides various health resources [7]. In reality, media can be a vital health educator, bridging the information gap between health care practitioners and the public in cancer communication [8,9]. The media publicize activities or campaigns from medical or governmental sectors and translate medical language and other jargon into public language [10,11]. Nowadays, besides health professionals, the internet has become a popular source of health information [12]. It changes the way an individual makes health decision, perceives health issues, and interacts with health professionals [13]. Cancer-related topics are the most commonly searched health topics online [14], and cancer news coverage is one of the most consumed health news categories by online newsreaders [15]. The majority of online news sites contain specific sections for conveying health news coverage: it allows information-oriented news consumers and active health information seekers to obtain timely health information about issues that concern them [16,17].

The health belief model explains that individual intentions and motivation of health behaviors are mostly subject to external cues, such as interpersonal communications and media consumptions [18]. It was defined as a variable called cue to action [19]. Based on this viewpoint, Tanner and Friedman [20] linked the concept of mobilizing information (MI) onto the theoretical ground of the health belief model. Originally, MI was a type of information widely examined by scholars from political science and journalism studies [21,22]. By definition, in the realm of public health, MI refers to a type of mediated information that could call individuals to carry out particular health actions (eg, quit smoking) or trigger them to act on preexisting attitudes effectively. Preexisting attitudes involve one's overall evaluation on a subject matter which later forms preconceived notions even if they did not experience it personally [23]. Since health information from media channels is usually persuasive, it acts as an actuator for health behavior change [7]: when individuals encounter persuasive health information, especially MI, their cognition and memories related to the particular health issue would be evocated and, in turn, brought to further actions. Specifically, MI includes the venue, time, and contact information of a health campaign, detailed educational information of how to a prevent certain disease or reduce risks, and additional links to provide further readings or interactions [20,24]. The individuals could actively or passively engage with different types of health news content that contain a certain type of MI. Thus, MI can reach newsreaders easily during their routine news consumption.

Several studies have examined MI in different health news coverage, but previous researchers found that there was very limited MI presented in the news content. For example, one study found that there was only approximately 30% of chronic disease coverage presented MI in Canadian aboriginal newspapers [25]. Another study reported that only 3 in 10 online health news articles from American local television websites presented MI [20]. Moreover, previous studies treated MI as an affiliated coding variable, scrutinizing on the present rate or types of MI instead of investigating the in-depth relationships between MI and other news components, such as news source and news focus. Also, most of the past studies examining MI are from the Western and English media channels [26,27]; little is known about how the Eastern and non-English media provide MI in health news coverage.

In Malaysia, media channels practice multilingual media policy to serve different communities [28]; the characteristics of the local media industry differ from other countries. Even in the local Chinese society, the media preference varies based on the socioeconomic background or other factors [28]. Thus, it is research-worthy for analyzing and comparing the local media contents published in different languages. Scholars reported that the way the media cover cancer news depends highly on who the target audience is [29]. Therefore, the focal points of the approach of presenting MI on cancer issues might differ from the local English media and the Chinese media. Since Malaysian Chinese populations had the highest cancer incidence rate among 3 main ethnicities [5], it is vital to know how Malaysian English and Chinese online news sites cover cancer issues and how these news sites mobilize individuals to prevent cancer. To fill the abovementioned research gaps, this study aims to answer the following research questions:

- RQ1a: What are the present rates, patterns, and types of MI in cancer news coverage on an English news site and a Chinese news site?
- RQ1b: Are there differences in the present rates, patterns, and types of MI on cancer coverage between an English news site and a Chinese news site?
- RQ2a: What are the characteristics of news components (ie, cancer type, news sources, news focus, and cancer risk factors) in cancer news coverage that provided MI on an English news site and a Chinese news site?
- RQ2b: Are there differences in the characteristics of news components (ie, cancer type, news sources, news focus, and cancer risk factors) in cancer news coverage that provided MI on an English news site and a Chinese news site?
- RQ3: Are there associations in the cancer coverage presenting MI between news sources, news focus, and cancer risk factors on an English news site and a Chinese news site?

Methods

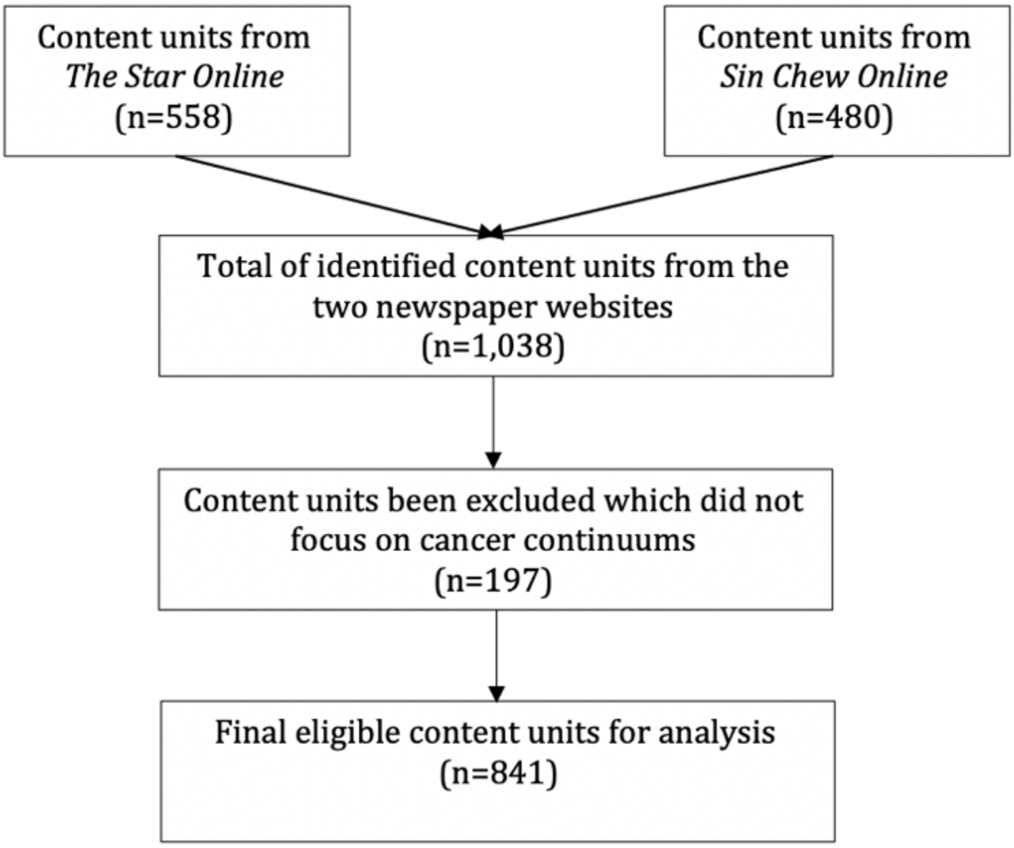
In this study, we applied quantitative content analysis to examine cancer news articles that contain MI on selected Malaysian English and Chinese newspapers with online versions and compare the differences between the newspaper websites.

Selection of Content Units

According to the Reuters Institute Digital News Report 2019 [30], The Star Online occupied 29% of the local weekly online news media use, which ranked the highest among all English newspaper websites. Sin Chew Online occupied 12% of it, which ranked the highest score among all local Chinese newspaper websites.

The unit of analysis of this study was the news article. News articles were retrieved using a keyword search of the homepages of individual newspaper websites. The keywords such as “cancer” OR “cancer prevention” OR “cancer treatment” OR “breast cancer” OR “lung cancer” OR “prostate cancer” OR “colon cancer,” etc, appearing in the headlines, subheadlines, first 3 paragraphs and lead paragraphs were used to find articles about MI on cancer prevention. We excluded news articles that covered obituaries of public figures who died from cancer and did not focus on cancer treatment or suggestions for cancer prevention. A total of 841 news articles were identified from 2017 to 2019. The process of selecting and excluding news articles for analysis is shown in Figure 1.

Figure 1. Process of selecting and excluding content units.



Research Instrument and Descriptions

Interrelated coding book and coding sheet were predesigned for the data collection purpose. There are 5 variables in the coding instruments that were conceptualized based on previous studies. There are several options under each variable; the number of options that should be chosen depends on the measurement and operationalization of each variable. All the variables were identified as nominal measurements. The operational definitions of each coding variable are shown in Multimedia Appendix 1.

Description of Cancer Type

Two categories of cancers (Multimedia Appendix 2) were examined in this study (ie, highly preventable cancers and highly detectable cancers). This description was adopted from Moriarty and Stryker's categorization of cancers [31], which was approved by the American Cancer Society. If an article covered any cancer-related topic without focusing on a specific type of cancer or covered more than one particular type of cancer, such as coverage on cancer care facilities' improvement, cancer-related health policy change, or statistic report on incidence rate/mortality rate, the article was then coded as cancer in general. Coders were requested to identify one type of cancer for a news article only.

Description of News Source

News source has a vital impact on health news coverage, and it is also a crucial variable for content analysis [32]. The description of news source in this study was adopted from past studies [33,34]. Six types of news sources were examined: medical journal, research institution, pharmaceutical company, government agency, nongovernmental organization (NGO), and other. Coders coded more than one news source for an article if applicable.

Description of News Focus

Different news articles cover an issue through different foci. The description of news focus was based on Tanner and Friedman's study [20]. There were 6 types of news focus under this study (ie, primary cancer prevention, secondary cancer prevention, medical treatment, social support/survivorship, medical research, and statistical report). Coders coded more than one type of news focus for an article if applicable.

Description of Cancer Risk Factors

This description aimed to identify the types of cancer risks mentioned in cancer news coverage. There were four categories of cancer risk factors that could be identified from cancer news articles, which are lifestyle risks, environmental/occupational risks, demographical risks, and medical risks. It was conceptualized and used in previous studies [35,36]. Coders coded more than one cancer risk factor for an article if applicable.

Description of Mobilizing Information

This description is the primary research focus of this study and includes 2 coding items (presence of MI and types of MI). First, the presence of MI in the news article was identified. Coders identified whether each article presented MI by coding present or absent in the coding sheet. If MI is absent from the article, the coders needed not to code the types of it.

If MI is presented in the article, then coders identified the types of MI in the article. There are 4 types of MI defined by previous research [20,21]: locational MI (ie, location of a hospital or event), identificational MI (ie, contact details of a health care provider), tactical MI (ie, suggestion or educational information regarding how to prevent cancer), and interactive MI (ie, hyperlinks for further readings). Coders coded more than one type for an article if applicable.

Intercoder Reliability

Two coders (the first author and a graduate student) randomly coded 10% of the identified news articles for intercoder reliability check. The graduate student received extensive training before coding the articles. We employed a Cohen kappa reliability test [37]. Based on the Altman strength of agreement, Cohen kappa value ranging from .61 to .80 indicates good agreement while values falling from .81 to 1.0 show very good agreement [38]. Our results demonstrated that some coding items obtained good agreement, such as tactical MI (k=.73, 95% CI .59-.87) and medical cancer risks (k=.74, 95% CI .59-.88), and the rest of items gained very good agreement between 2 coders, like cancer type (k=.89, 95% CI .78-.98) and presence of MI (k=.94, 95% CI .85-1.0).

Data Analysis

The obtained data were analyzed via SPSS (version 26.0, IBM Corp). First, we applied descriptive analysis and chi-square test to answer RQ1 and RQ2, including the frequencies, percentages, and differences of each news component. We then employed binary logistic regression (forward likelihood ratio method) to determine the associations between different news components and the presence of MI in online cancer news (RQ3). The level of significance in this study was set as $P<.05$.

Results

There were 585 out of 841 cancer news articles on The Star Online and Sin Chew Online provided at least one type of MI in a span of 36 months from January 2017 to December 2019 (Table 1). The overall present rate of MI was 69.6% (585/841), which means nearly 7 in 10 of cancer news articles provided MI. Specifically, 64.4% (283/436) of articles on The Star Online, and 74.6% (302/405) of news articles on Sin Chew Online provided MI. The present rate of MI on Sin Chew Online was higher than The Star Online (χ^2=9.25, df=1; P=.003).

Table 1. The presence of mobilizing information in cancer news coverage on The Star Online and Sin Chew Online (2017-2019; n=841).

MI[a] presence	The Star Online, n (%)	Sin Chew Online, n (%)	Overall, n (%)
Present	283 (64.4)	302 (74.6)	585 (69.6)
Absent	152 (35.6)	103 (25.4)	258 (30.4)
Total	436 (100)	405 (100)	841 (100)

[a] MI: mobilizing information.

The overall patterns of monthly distribution showed a fluctuation in both newspaper websites (Figures 2 and 3). The patterns of cancer news articles providing MI between the 2 news sites were slightly different. For The Star Online, the peak was in October 2019, and there were 18 cancer news articles contained MI. However, no cancer news article provided MI in July 2019. Sin Chew Online had 2 peak months, October 2017 and September 2019. There were 18 cancer news articles delivered MI in both months. The lowest month for Sin Chew Online was April 2017, which only had 2 cancer news articles that provided MI.

Figure 2. Numbers of cancer news articles provided mobilizing information on The Star Online from Jan 2017 to Dec 2019.

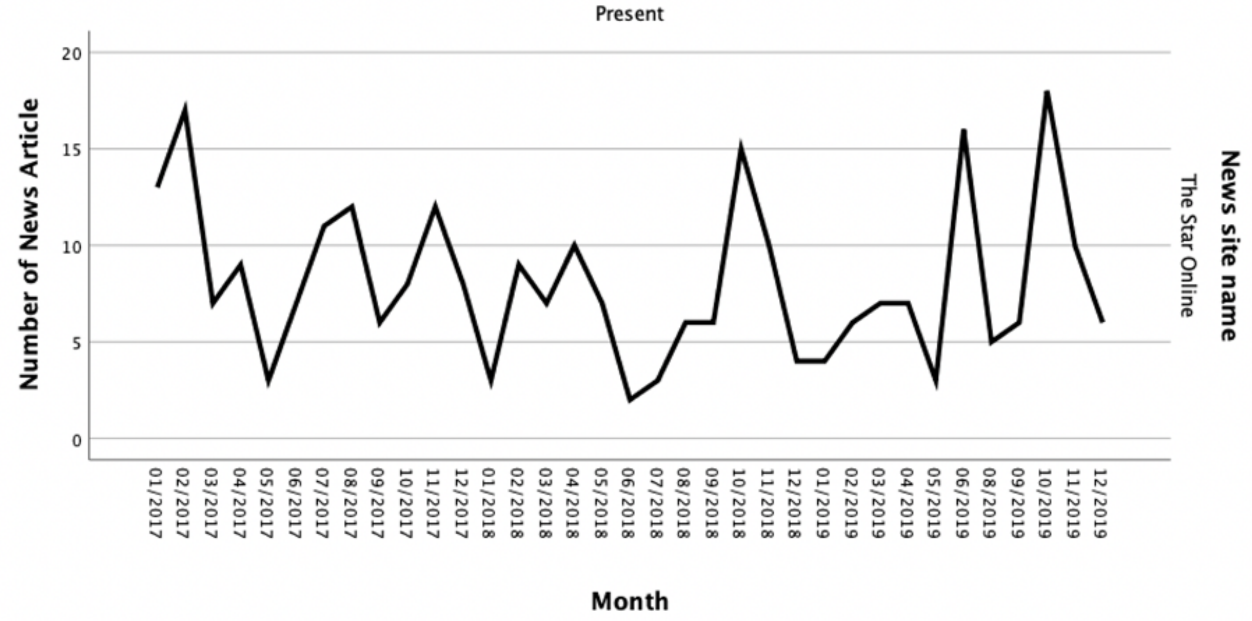
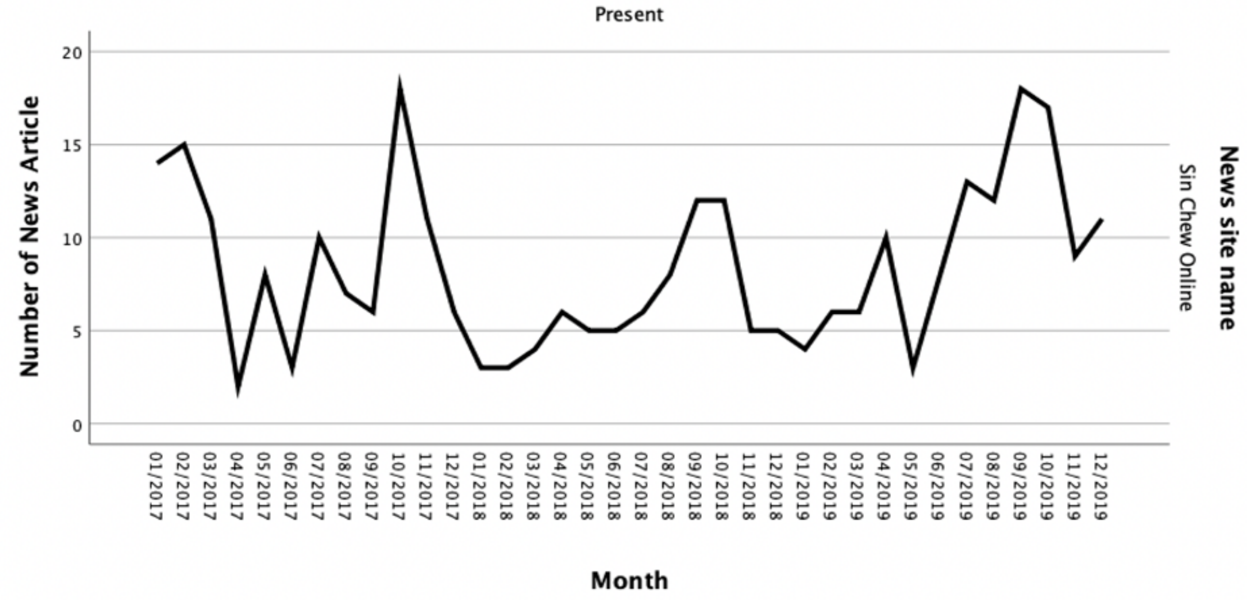


Figure 3. Numbers of cancer news articles provided mobilizing information on Sin Chew Online from Jan 2017 to Dec 2019.



For the types of MI (Table 2), the majority of news articles (419/585, 71.6%) from both news sites provided tactical MI. Locational MI was the second most frequently presented in the news articles, which occupied 25.1% (109/585). Besides, there were similar patterns on providing identificational MI (96/585, 16.4%) and interactive MI (107/585, 18.3%), both provided by less than 20% of cancer news articles. The chi-square results indicated that news articles in The Star Online provided more interactive MI (χ^2_1=12.08, *P*=.001), while news articles in Sin Chew Online provided more identificational MI (χ^2_1=5.44, *P*=.03).

Table 2. Distributions of types of mobilizing information (n=585).

Type of MI[a]	The Star Online, n (%)	Sin Chew Online, n (%)	Overall, n (%)	χ^2	*P* value
Locational MI	91 (15.6)	109 (18.6)	200 (34.2)	1.0	.34
Identificational MI	36 (6.1)	60 (10.3)	96 (16.4)	5.4	.03
Tactical MI	192 (32.8)	227 (38.8)	419 (71.6)	3.9	.05
Interactive MI	68 (11.6)	39 (6.7)	107 (18.3)	12.1	.001

[a] MI: mobilizing information.

Descriptive analysis and chi-square test for each news component were conducted to answer RQ2a and RQ2b (Table 3). More than half of the cancer news articles that provided MI did not focus on a specific cancer (313/585, 53.5%), and there were overall 36.1% (211/585) of news articles focused on highly detectable cancers such as breast cancer, prostate cancer, liver cancer, and stomach cancer. Only 10.4% (61/585) of articles gave attention to highly preventable cancers such as lung cancer and cervical cancer. There was a significant difference between the websites in covering different cancer types (χ^2_1=8.05, *P*=.02).

Pertaining to news sources, more than half (333/585, 57%) of the cancer news articles that provided MI from both newspaper websites interviewed staff from medical institutions, followed by NGOs (208/585, 35.5%). Less than 20% of the news articles interviewed cancer patients, cancer survivors, or other individuals who did not belong to any institution (102/585, 17.5%) or government agency (97/585, 16.6%). Very few news articles cited medical journals (37/585, 4.7%) or interviewed pharmaceutical companies (8/585, 1%). The chi-square results indicated that there was a significant difference between newspaper websites and news sources used while reporting cancer news. (11.3% [66/585] vs 5.3% [31/585]; χ^2_1=12.05, *P*=.001). Specifically, Sin Chew Online skewed to refer to government agencies as the source in reporting cancer news.

As for news focus, the majority (410/585, 70%) of cancer news articles that provided MI focused on primary cancer prevention and almost half (285/585, 48.7%) focused on secondary cancer prevention. In addition, 28.0% news articles that focused on social support contained MI (164/585), followed by medical treatment (127/585, 21.7%), medical research (80/585, 13.7%), and statistical reports on cancer issues (69/585, 11.8%). The chi-square results showed that news articles in The Star Online were more likely to focus on medical treatment (13% [76/585] vs 8.7% [51/585]; χ^2_1=8.54, *P*=.004); news articles in Sin Chew Online were more likely to look at primary cancer prevention (39.3% [230/585] vs 30.8% [180/585]; χ^2_1=10.98, *P*=.001).

In terms of cancer risk factors, 63.0% (368/585) of cancer news articles that provided MI mentioned demographical risks, followed by medical risks (281/585, 48.1%) and lifestyle risks (236/585, 40.4%), respectively. Only 3.3% (19/585) of news articles gave attention to environmental/occupational risks that may cause cancer. The chi-square result showed that news articles in Sin Chew Online focused more on lifestyle risks than articles in The Star Online (23.6% [138/585] vs 16.8% [98/585]; χ^2_1=7.43, *P*=.007).

Table 3. Descriptive analysis of news components in cancer news coverage that provided mobilizing information (n=585).

Variable and category	The Star Online, n (%)	Sin Chew Online, n (%)	Overall, n (%)	χ^2	*P* value
Cancer type	—[a]	—	—	8.1	.02
General cancer topic	139 (23.8)	174 (29.7)	313 (53.5)	—	—
Highly preventable cancer	39 (6.7)	22 (3.8)	61 (10.5)	—	—
Highly detectable cancer	105 (17.9)	106 (18.1)	211 (36.0)	—	—
News source	—	—	—	—	—
Medical journal	21 (3.6)	16 (2.7)	37 (6.3)	1.1	.31
Medical institution	173 (29.6)	160 (27.4)	333 (57.0)	4.0	.06
Pharmaceutical company	5 (0.9)	3 (0.5)	8 (1.4)	0.7	.50
Government agency	31 (5.3)	66 (11.3)	97 (16.6)	12.1	.001
Nongovernmental organization	105 (17.9)	103 (17.6)	208 (35.5)	0.6	.49
Other individuals	53 (9.1)	49 (8.4)	102 (17.5)	0.4	.45
News focus	—	—	—	—	—
Primary cancer prevention	180 (30.8)	230 (39.3)	410 (70.0)	11.0	.001
Secondary cancer prevention	149 (25.5)	136 (23.2)	285 (48.7)	3.4	.07
Medical treatment	76 (13.03)	51(8.7)	127 (21.7)	8.5	.004
Social support	77 (13.2)	87 (14.8)	164 (28.0)	0.2	.71
Medical research	47 (8.0)	33 (5.6)	80 (13.6)	4.0	.05
Statistical report	34 (5.8)	35 (6.0)	69 (11.8)	0.03	.09
Cancer risk factors	—	—	—	—	—
Lifestyle risks	98 (16.8)	138 (23.6)	236 (40.4)	7.4	.007
Environmental/occupational risks	8 (1.4)	11 (1.9)	19 (3.3)	0.3	.65
Demographic risks	184 (31.5)	184 (31.5)	368 (63.0)	1.1	.35
Medical risks	142 (24.3)	139 (23.8)	281 (48.1)	1.0	.32

[a] Not applicable.

A binary logistic regression analysis (forward likelihood ratio method) was conducted to answer RQ3. Table 4 presents these results accordingly. First, the results indicated that news articles that interviewed NGOs and medical institutions were more likely to present MI (OR 2.77, 95% CI 1.89-4.05; OR 1.85, 95% CI 1.33-2.58, respectively). News articles that cited medical journals as the source were also more likely to present MI (OR 2.28, 95% CI 1.03-5.01), but the significant level was lower than NGOs and medical institutions (95% [*P*=.04] vs 99% [*P*=.001]). Therefore, NGOs, being the news source, obtained the strongest association with the presence of MI. There was no significant association found for other news sources, such as medical journals, pharmaceutical companies, or governmental agencies in the examined news articles.

Second, for news focus, the binary logistic regression results reported that news articles were more likely to present MI when they focused on primary cancer prevention (OR 97.70, 95% CI 46.97-203.24) and secondary cancer prevention (OR 22.12, 95% CI 12.16-40.22). However, news articles that focused on medical research were less likely to present MI (OR .33, 95% CI .12-.37). The equation model shows that articles that focus on primary cancer prevention obtained the strongest association with the presence of MI. Other news focus such as medical treatment, social support/survivorship, or statistical report were not significantly associated with the presence of MI.

In terms of cancer risk factors, the binary logistic regression results showed that all 4 types of cancer risk factors (ie, lifestyle, environmental/occupational, demographic, and medical) were positively associated with the presence of MI in the online cancer news. However, the significant levels varied. The news articles that reported lifestyle risks obtained the strongest association with the presence of MI (OR 186.28; 95% CI 44.83-773.96), followed by demographic risks (OR 8.97; 95% CI 6.08-12.25), and medical risks (OR 3.07; 95% CI 2.08-4.53). The news articles that reported environmental/occupational risks were also associated with the presence of MI (OR 3.83; 95% CI 1.15-12.76). However, when comparing to the other 3 risk categories, the significance level of the association between environmental/occupational risks and the presence of MI was relatively weak (only at 95%).

Table 4. Associations between news sources, news focus, cancer risk factors, and the presence of mobilizing information (n=841).

Variable and step	Category	OR[a] (95% CI)	*P* value
News source[b]			
1	Nongovernmental organization	1.97 (1.40-2.77)	.001
2	Nongovernmental organization	2.69 (1.84-3.93)	.001
	Medical institution	1.91 (1.37-2.67)	.001
3	Nongovernmental organization	2.77 (1.89-4.05)	.001
	Medical institution	1.85 (1.33-2.58)	.001
	Medical journal	2.28 (1.03-5.01)	.04
News focus[c]			
1	Primary cancer prevention	64.30 (312.31-129.97)	.001
2	Primary cancer prevention	91.05 (44.55-186.09)	.001
	Secondary cancer prevention	23.32 (12.96-41.95)	.001
3	Primary cancer prevention	97.70 (46.97-203.24)	.001
	Secondary cancer prevention	22.12 (12.16-40.22)	.001
	Medical research	.33 (0.12-0.37)	.001
Cancer risk factors[d]			
1	Lifestyle risks	85.88 (21.15-384.65)	.001
2	Lifestyle risks	138.12 (33.63-567.37)	.001
	Demographic risks	8.52 (5.88-12.35)	.001
3	Lifestyle risks	180.50 (43.54-748.36)	.001
	Demographic risks	8.50 (5.80-12.49)	.001
	Medical risks	2.94 (2.00-4.32)	.001
4	Lifestyle risks	186.28 (44.83-773.96)	.001
	Demographic risks	8.97 (6.08-12.25)	.001
	Medical risks	3.07 (2.08-4.53)	.001
	Environmental/occupational risks	3.83 (1.15-12.76)	.03

[a] OR: odds ratio.

[b] The −2 log likelihood = 997.94, Cox & Snell R^2=.04, Nagelkerke R^2=.06.

[c] The −2 log likelihood = 476.57, Cox & Snell R^2=.48, Nagelkerke R^2=.69.

[d] The −2 log likelihood = 659.42, Cox & Snell R^2=.36, Nagelkerke R^2=.51.

Discussion

Principal Findings

As the media has been instrumental in creating public awareness about cancer and chronic disease prevention, understanding the media's role in mobilizing the public regarding cancer prevention is of utmost importance. This study provides the first comprehensive and comparative assessment of MI in cancer news on the selected Malaysian English and Chinese newspapers with online versions. Of importance, our results show that there were nearly 70% of the online cancer news articles in The Star Online and Sin Chew Online presented at least one type of MI. It indicates that the selected Malaysian newspaper websites have an awareness of mobilizing and calling health behavior change for its readers. However, our findings are contrary to previous studies that examined MI in cancer coverage in Western countries [22,24-26]. Several possible explanations could support our findings. First, since the past studies that examined MI in cancer news coverage were conducted at least 10 years ago, the patterns of cancer news coverage have already changed, especially for online news media [22,24-26]. Second, Hoffman [20] introduced another type of MI called interactive MI, which is only applicable for online news content. This could be one of the reasons why online cancer news contained more MI as compared to print media. Third, the increasing rates of cancer incident and mortality around the world [1], the seriousness of the disease, health information needs among the people, and the media attention of cancer issues that differ from the past have become another dependable rationale to support our results. Fourth, culture and social background would give impacts to health coverage [39]. Thus, our results report that the characteristics of cancer news coverage in the selected Malaysian newspaper websites were quite different from the

Western countries. Of note, we also found that cancer news articles in Sin Chew Online were more likely to present MI than those in The Star Online. This may due to the increased cancer incident rate reported among the Chinese population [4,5], and Sin Chew Online realized that it is crucial to promote and mobilize cancer prevention to its targeted readers.

Our results illustrated that there was an unbalanced monthly distribution in terms of the number of cancer news articles that provided MI. In certain months, there were more cancer news articles found providing MI, such as in February, October, and November. This may due to the fact that World Cancer Day falls in February, Children's Cancer Awareness Month falls in September, and Breast Cancer Awareness Month falls in October. These occasions gained more media attention because of shared journalistic values; journalists usually cover health topics according to the allocation of relevant campaigns or events [40]. In the same vein, Varga and colleagues [40] reported that the media coverage on certain cancers mushroomed during their awareness months, such as breast cancer and prostate cancer, and especially on social media platforms. However, in certain months, there were fewer cancer news articles reporting MI, such as in June and July. These months are the busier months throughout the year as there are many local festivals and school holidays happen in these months. The Hari Raya Aidilfitri, Gawai Dayak, King's birthday, and Dragon Boat Festival happen in these 2 months. Negative news such as reporting cancer issues and related death tolls would likely give way to positive news and announcements such as big sales and discounts, tourism, and Open House. Hence, the news distribution of MI on cancer prevention was found less in June and July. Speaking of the types of MI, we found that the majority of cancer news provided tactical MI, which contained how-to health information and detailed suggestions about cancer prevention. Such news provides a clear picture to the readers of how to prevent cancers in their daily life.

We found several differences between The Star Online and Sin Chew Online regarding news components providing MI in online cancer news. First, the results indicated that there was a significant difference in terms of cancer types portrayed in the 2 studied online news websites. More than half of cancer news that provided MI focused more than one specific type of cancer. It mainly covered health campaigns or events related to cancer prevention, followed by highly detectable cancer and highly preventable cancer. This phenomenon is consistent with the previous study, which applied the same categorization of cancer types [31]. By comparison, cancer news from Sin Chew Online focused more on general cancer topics while The Star Online is more likely to cover highly preventable cancer issues.

For news source, the results showed that the majority of cancer coverage that provided MI interviewed professionals from medical institutions, such as public and private hospitals, universities or laboratories, followed by NGOs. This corroborates with previous studies that examined sourcing practices in various health coverage [41,42]. Interestingly, we found that cancer news articles that provided MI in The Star Online were more likely to cite medical journals from overseas countries, while those in Sin Chew Online were more likely to interview government agencies and NGOs from the local Chinese society, such as ancestral connective organizations and religious organizations serve for Chinese communities (ie, Chinese temples and Chinese-dominant churches). This shows that cancer coverage from different languages has varying source preferences subject to their journalistic values and social/cultural norms [39].

Pertaining to news focus, most cancer news providing MI focused on primary cancer prevention, such as local educational events or social marketing campaigns related to cancer issues, followed by secondary cancer prevention, like mobilizations on cancer screening for early detection and other medical interventions. Since the characteristics of MI include triggers and call the readiness among individuals, it is reasonable that the majority of cancer coverage that provided MI focused on prevention, rather than other topics like cancer treatment or cancer research [20]. We found that cancer news providing MI in The Star Online focused slightly more on topics related to secondary cancer prevention and medical treatment as compared to the coverage reported in Sin Chew Online, which was more likely to concentrate on primary cancer prevention and social support topics. This phenomenon reflects Malaysian media ecology when it comes to health news coverage.

As for cancer risk factors, our results showed different findings compared to the previous study where this description was adopted [35]. More than half of the examined news coverage mentioned demographic risks, like cancer-related with a specific gender or age. Most of these demographic risks pointed to female cancers, such as cervical cancer prevention, HPV vaccination, and breast cancer self-checking. Almost half of cancer news articles mentioned medical risks and lifestyle risks when providing MI. It corroborates the dominant characteristics of MI that promotes behavioral change on health issues [20], which advocates changes and adoptions of health behavior, as well as an understanding of medical knowledge and preventive strategies among the readers. However, it is worth noting that there was rarely cancer coverage that provided MI on environmental risks, such as natural resources pollutions, radiations, and occupational risks. The reason could be due to lack of knowledge on environmental issues and their association with cancer morbidity and mortality [43]. By comparing the two newspaper websites, we found that cancer news from Sin Chew Online was more likely to mention lifestyle risks when providing MI. It may subject to the cancer incidence in local Chinese communities; reducing risk lifestyle is one of the dominant approaches to avoid cancer occurrence [5,6].

The logistic regression analysis unveiled the in-depth findings of this study. This method is inspired by previous studies that used the same approach to analyze other media messages [34,44]. We examined the associations between 3 critical news components (ie, news sources, news focus, and cancer risk factors) and the presence of MI in online cancer news separately. For the news sources, our results illustrated that the cancer news articles were most likely to present MI when NGOs were interviewed as the source, followed by medical institutions and medical journals. This result agrees with a previous study finding that elite sources are the dominant news sources in health news coverage [32]. This study suggests that health journalists should be more active in engaging with NGOs and medical institutions. On this count, NGOs could bring up-to-date resources for events and campaigns regarding cancer prevention, and medical institutions can provide credible and applicable health information for cancer prevention.

A finding of interest was that MI was more likely to be presented in online cancer news when articles focused on primary cancer prevention, such as health education, consultation, health events or campaigns, followed by secondary cancer prevention, like promotions of early detection, screening, or diagnosis. On the other hand, MI was less likely to be presented when articles covered topics related to medical research. Thus, we recommend that health journalists should devote more attention on how to cover topics related to cancer education, local events, and campaigns, which would mobilize and persuade the newsreaders to take proactive actions to prevent cancer instead of merely providing medical research-related news retrieved from collaborative news agencies.

Regarding cancer risk factors, we found that even though every type of cancer risk was positively associated with the presence of MI, the probability level was different. The articles that reported lifestyle risks such as obesity, alcohol consumption, or smoking causing cancer were positively associated with the presence of MI in online cancer news. This is consistent with previous findings which showed that since various cancers occurring today are related to unhealthy lifestyles, cancer news coverage switched their major attention from highlighting environmental causes to discussing lifestyle causes [35]. Our results also showed that the mentions of demographic risks, like risks related to gender and age, were positively associated with the presence of MI. However, we found out that the majority of news articles that mentioned demographic risks pointed to females, which suggests that future events or campaigns could focus on mobilizing females to prevent breast and cervical cancer, as well as educating this target group to raise their awareness on cancers that most prevalent among females. There were limited news articles on demographic risks provided MI that targeted males even though prostate cancer and testicular cancer are two serious health issues for males [45]. We suggest that health journalists need to notice that cancer coverage should not skew to cancers related to females only. It is also essential to raise awareness and mobilize males to prevent cancers that are associated with them. The lack of coverage on cancers related to males calls for the need for research on effective strategies to reduce health information disparities among the male population.

Limitations

This study contains some limitations. First, we only applied a keyword searching strategy for searching online news articles, and we might have missed some relevant news articles due to human errors. Future researchers should improve news sample selecting strategy by applying technology-based approaches to minimize human errors. Second, we only analyzed MI in online cancer news from the selected Malaysian newspaper websites. We did not consider the cancer news published in Malay and Tamil newspaper websites, which are two other major languages spoken in Malaysia. The representativeness of this study, hence, is limited. More studies should give attention to cancer news published in Malay and Tamil; it would draw a more comprehensive picture in terms of covering cancer issues and providing MI in Malaysian media. Third, we determined the original definition of tactical MI is ambiguous, which points to “the explicit and implicit instructions for certain behavior” [19,20]. The understanding and operationalization of this concept highly depend on different researchers and newsreaders. Thus, communication scholars need to argue and re-theorize this definition in future studies. Fourth, it is worth noting that there is a lot of variation in the products and content of the agencies used as sources, such as their stature, professional staff, mission, and other characteristics that would explain or influence the variance in data collected. In this study, the researchers only categorized the sources in generic forms. Future studies could apply textual analysis, examining the latent content of news coverage or case study to explore these variations in detail. Fifth, this study did not compare the differences or correlations of representation of MI on the internet, print, and broadcast media. Hence, future research could examine whether if different types of MI are presented differently on different mediums. Last, our results only could represent the way of presenting MI by the online news media, but the effects of MI toward individuals are unknown. Future studies should examine the effects of MI and behavioral intention or outcomes among newsreaders.

Conclusion

This study is one of the first scholarly attempts to throw light on analyzing and comparing MI of cancer news on selected English and Chinese newspaper websites. The findings and discussions corroborate new practical, theoretical, and methodological knowledge in health communication and journalism that would help public health researchers, medical professionals, health journalists, health policymakers, and NGOs in their researches and decision makings. The comprehensive picture of this study shows the characteristics and in-depth relationships between MI and different news components. For future studies, researchers could focus on online cancer news in other languages and also compare the differences between those published in native/national language and English. Also, experimental designs should be developed to examine the effects of MI on cancer prevention among different individuals. Furthermore, our methodological use could be applied in future studies that examine other variables in health news coverage, such as frame building and supplemental information.

Acknowledgments

This study is part of the first author’s dissertation and is not funded by any internal or external grant.

Conflicts of Interest

Not declared.

Multimedia Appendix 1
Definitions of coding items.
[DOCX File , 20 KB - jmir_v23i6e26019_app1.docx]

Multimedia Appendix 2
Classifications of cancer type.
[DOCX File , 15 KB - jmir_v23i6e26019_app2.docx]

References

1. Cancer overview. Geneva: World Health Organization; 2020 Jan. URL: https://www.who.int/health-topics/cancer#tab=tab_1 [accessed 2020-05-15]
2. Lim GCC. Overview of cancer in Malaysia. Jpn J Clin Oncol 2002 Feb 01;32(suppl 1):S37-S42. [doi: 10.1093/jjco/hye132]
3. Cancer today—Malaysia. International Agency for Research on Cancer. 2021 Mar. URL: https://gco.iarc.fr/today/data/factsheets/populations/458-malaysia-fact-sheets.pdf [accessed 2021-06-01]
4. Malaysia National Cancer Registry Report 2012-2016. Malaysia Ministry of Health. 2019. URL: https://www.moh.gov.my/moh/resources/Penerbitan/Laporan/Umum/2012-2016%20(MNCRR)/MNCR_2012-2016_FINAL_(PUBLISHED_2019).pdf [accessed 2021-06-01]
5. [Report: Cancer incidence rates are increasing among Malaysians; Chinese ranked the highest]. Sin Chew Online. 2020 Jan. URL: https://www.sinchew.com.my/content/content_2190490.html [accessed 2020-05-15]
6. Colditz G, Hunter D. Cancer Prevention: The Causes and Prevention of Cancer. Volume 1. Amsterdam: Springer; 2000.
7. Dutta M. Content effects: personal and public health. In: Rössler P, editor. The International Encyclopedia of Media Effects. Hoboken: Wiley-Blackwell; 2017.
8. Sern TJ, Firdaus AS, Zanuddin H. Government initiatives and public awareness on HIV/AIDS in Malaysia: where are we after 30 years? Malay J Commun 2016 Dec 1;32(2):545-565. [doi: 10.17576/jkmjc-2016-3202-27]
9. Miyawaki R, Shibata A, Ishii K, Oka K. News coverage of cancer in Japanese newspapers: a content analysis. Health Commun 2017 Apr;32(4):420-426. [doi: 10.1080/10410236.2016.1138391] [Medline: 27315217]
10. McLuhan M. Understanding Media: The Extension of Man. Berkeley: Gingko Press; 2003.
11. Schiavo R. Health Communication: From Theory to Practice. 2nd Edition. San Francisco: Jossey-Bass; 2014.
12. Nguyen GT, Bellamy SL. Cancer information seeking preferences and experiences: disparities between Asian Americans and Whites in the Health Information National Trends Survey (HINTS). J Health Commun 2006 Feb;11 Suppl 1(sup001):173-180. [doi: 10.1080/10810730600639620] [Medline: 16641082]
13. Jiang S. How does online patient-provider communication heal? Examining the role of patient satisfaction and communication experience in China. Health Commun 2019 Nov 10;34(13):1637-1644. [doi: 10.1080/10410236.2018.1517634] [Medline: 30198772]
14. Shim M. Connecting internet use with gaps in cancer knowledge. Health Commun 2008 Sep 08;23(5):448-461. [doi: 10.1080/10410230802342143] [Medline: 18850392]
15. Hurley RJ, Riles JM, Sangalang A. Online cancer news: trends regarding article types, specific cancers, and the cancer continuum. Health Commun 2014 Jan 28;29(1):41-50. [doi: 10.1080/10410236.2012.715538] [Medline: 23356571]
16. Dutta-Bergman MJ. Primary sources of health information: comparisons in the domain of health attitudes, health cognitions, and health behaviors. Health Communication 2004 Jul;16(3):273-288. [doi: 10.1207/s15327027hc1603_1]
17. Shim M, Kelly B, Hornik R. Cancer information scanning and seeking behavior is associated with knowledge, lifestyle choices, and screening. J Health Commun 2006 Feb;11 Suppl 1(sup001):157-172. [doi: 10.1080/10810730600637475] [Medline: 16641081]
18. Jones CL, Jensen JD, Scherr CL, Brown NR, Christy K, Weaver J. The Health Belief Model as an explanatory framework in communication research: exploring parallel, serial, and moderated mediation. Health Commun 2015 Jul 10;30(6):566-576 [FREE Full text] [doi: 10.1080/10410236.2013.873363] [Medline: 25010519]
19. Rosenstock IM. The health belief model and preventive health behavior. Health Educ Behav 1974 Dec 01;2(4):354-386. [doi: 10.1177/109019817400200405]
20. Tanner A, Friedman DB. Health on the Web: an examination of health content and mobilising information on local television Websites. Inform Health Soc Care 2011 Jan 21;36(1):50-61. [doi: 10.3109/17538157.2010.505311] [Medline: 20964501]
21. Hoffman LH. Is internet content different after all? A content analysis of mobilizing information in online and print newspapers. Journ Mass Commun Q 2016 Jun 25;83(1):58-76. [doi: 10.1177/107769900608300105]
22. Shaker L. Community newspapers play significant role in election. Newsp Res J 2011 Jan 01;32(1):6-18. [doi: 10.1177/073953291103200102]

23. Mazaheri E, Basil DZ, Yanamandram V, Daroczi Z. The impact of pre-existing attitude and conflict management style on customer satisfaction with service recovery. J Retail Consumer Serv 2011 May;18(3):235-245. [doi: 10.1016/j.jretconser.2010.11.004]
24. Cohen EL, Caburnay CA, Luke DA, Rodgers S, Cameron GT, Kreuter MW. Cancer coverage in general-audience and Black newspapers. Health Commun 2008 Sep 08;23(5):427-435. [doi: 10.1080/10410230802342176] [Medline: 18850390]
25. Hoffman-Goetz L, Shannon C, Clarke J. Chronic disease coverage in Canadian aboriginal newspapers. J Health Commun 2003 Sep;8(5):475-488. [doi: 10.1080/10810730390233299] [Medline: 14530149]
26. Friedman DB, Hoffman-Goetz L. Cancer coverage in North American publications targeting seniors. J Cancer Educ 2003 Mar 1;18(1):43-47. [doi: 10.1207/s15430154jce1801_14] [Medline: 12825634]
27. Hoffman-Goetz L, Friedman DB. Disparities in the coverage of cancer information in ethnic minority and mainstream mass print media. Ethn Dis 2005;15(2):332-340. [Medline: 15825981]
28. Yang LF, Ishak MSA. Framing inter-religious dispute: a comparative analysis of Chinese-, English-, and Malay-language newspapers in Peninsular Malaysia. Chin J Commun 2015 Nov 02;9(2):103-123. [doi: 10.1080/17544750.2015.1096802]
29. King AJ, Jensen JD, Guntzviller LM, Perez Torres D, Krakow M. Ethnic newspapers and low-income Spanish-speaking adults: influence of news consumption and health motivation on cancer prevention behaviors. Ethn Health 2018 May 24;23(4):410-424 [FREE Full text] [doi: 10.1080/13557858.2017.1280133] [Medline: 28116925]
30. Digital News Report: Malaysia 2019. Reuters Institute for the Study of Journalism. 2020 Feb. URL: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf [accessed 2020-05-15]
31. Moriarty CM, Stryker JE. Prevention and screening efficacy messages in newspaper accounts of cancer. Health Educ Res 2008 Jun 01;23(3):487-498. [doi: 10.1093/her/cyl163] [Medline: 17289658]
32. De Dobbelaer R, Van Leuven S, Raeymaeckers K. The human face of health news: a multi-method analysis of sourcing practices in health-related news in belgian magazines. Health Commun 2018 May 03;33(5):611-619. [doi: 10.1080/10410236.2017.1287237] [Medline: 28278607]
33. Holton A, Weberling B, Clarke CE, Smith MJ. The blame frame: media attribution of culpability about the MMR-autism vaccination scare. Health Commun 2012 Oct;27(7):690-701. [doi: 10.1080/10410236.2011.633158] [Medline: 22236220]
34. Moriarty CM, Jensen JD, Stryker JE. Frequently cited sources in cancer news coverage: a content analysis examining the relationship between cancer news content and source citation. Cancer Causes Control 2010 Jan 26;21(1):41-49. [doi: 10.1007/s10552-009-9432-x] [Medline: 19784789]
35. Jensen JD, Moriarty CM, Hurley RJ, Stryker JE. Making sense of cancer news coverage trends: a comparison of three comprehensive content analyses. J Health Commun 2010 Mar;15(2):136-151. [doi: 10.1080/10810730903528025] [Medline: 20390983]
36. Min HS, Yun EH, Park J, Kim YA. Cancer news coverage in Korean newspapers: an analytic study in terms of cancer awareness. J Prev Med Public Health 2020 Mar;53(2):126-134 [FREE Full text] [doi: 10.3961/jpmph.19.256] [Medline: 32268467]
37. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
38. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall; 1991.
39. Tang L, Peng W. Culture and health reporting: a comparative content analysis of newspapers in the United States and China. J Health Commun 2015 Nov 20;20(2):187-195. [doi: 10.1080/10810730.2014.920060] [Medline: 25411911]
40. Vraga EK, Stefanidis A, Lamprianidis G, Croitoru A, Crooks AT, Delamater PL, et al. Cancer and social media: a comparison of traffic about breast cancer, prostate cancer, and other reproductive cancers on Twitter and Instagram. J Health Commun 2018 Jan 09;23(2):181-189. [doi: 10.1080/10810730.2017.1421730] [Medline: 29313761]
41. De Dobbelaer R, Van Leuven S, Raeymaeckers K. Dirty dancing: health journalists and the pharmaceutical industry a multi-method study on the impact of pharma PR on magazine health news. Public Relations Rev 2017 Jun;43(2):450-459. [doi: 10.1016/j.pubrev.2017.02.002]
42. Gabore SM. Western and Chinese media representation of Africa in COVID-19 news coverage. Asian J Commun 2020 Aug 05;30(5):299-316. [doi: 10.1080/01292986.2020.1801781]
43. Nik Hasan NN. The Representation of Environmental News: A Comparative Study of the Malaysian and New Zealand press [Thesis]. Christchurch: University of Canterbury; 2007.
44. Liu Y, Tsai J, Chen Y. Beyond altruism: framing organ donation in a 19-year review of Chinese news coverage. J Health Commun 2019 Nov 05;24(12):878-888. [doi: 10.1080/10810730.2019.1681564] [Medline: 31687890]
45. Roy RK, Casson K. Attitudes toward testicular cancer and self-examination among Northern Irish males. Am J Mens Health 2017 Mar 21;11(2):253-261 [FREE Full text] [doi: 10.1177/1557988316668131] [Medline: 27645516]

Abbreviations

MI: mobilizing information
NGO: nongovernmental organization
OR: odds ratio

Edited by R Kukafka; submitted 25.11.20; peer-reviewed by L Reid-Chassiakos, A Rovetta, T Freeman; comments to author 02.02.21; revised version received 08.02.21; accepted 04.05.21; published 21.06.21.

Please cite as:
Zhang TH, Tham JS
Calls to Action (Mobilizing Information) on Cancer in Online News: Content Analysis
J Med Internet Res 2021;23(6):e26019
URL: https://www.jmir.org/2021/6/e26019
doi:10.2196/26019
PMID:34152283

Original Paper

Optimization of Primary Care Among Black Americans Using Patient Portals: Qualitative Study

Omar H Ordaz[1], BS; Raina L Croff[2], PhD; LaTroy D Robinson[1], BA; Steven A Shea[1], PhD; Nicole P Bowles[1], PhD

[1]Oregon Institute of Occupational Health Sciences, Oregon Health Sciences University, Portland, OR, United States
[2]Department of Neurology, School of Medicine, Oregon Health Sciences University, Portland, OR, United States

Corresponding Author:
Nicole P Bowles, PhD
Oregon Institute of Occupational Health Sciences
Oregon Health Sciences University
3181 SW Sam Jackson Park Rd L606
Portland, OR, 97239
United States
Phone: 1 503 494 2541
Email: bowlesn@ohsu.edu

Abstract

Background: Reduced patient portal use has previously been reported among Black Americans when compared with that of the general population. This statistic is concerning because portals have been shown to improve the control of chronic conditions that are more prevalent and severe in Black Americans. At their very simplest, portals allow patients to access their electronic health records and often provide tools for patients to interact with their own health information, treatment team members, and insurance companies. However, research suggests that Black American patients have greater concerns over a lack of support, loss of privacy, and reduced personalization of care compared with other Americans, which results in a disparity of portal use.

Objective: This qualitative investigation of primary care experiences of Black Americans from across the United States who participated in remote focus groups in April and May 2020 aims to explore the use and perceived value of patient portals to better understand any barriers to optimized treatment in the primary care setting.

Methods: We performed an inductive thematic analysis of 8 remote focus group interviews with 29 Black American patients aged 30-60 years to qualitatively assess the experiences of Black American patients with regular access to portals.

Results: Thematic analysis uncovered the following interrelated themes regarding patient portals in primary care: the optimization of care, patient empowerment, patient-provider communication, and patient burden.

Conclusions: In contrast to what has been described regarding the reluctance of Black Americans to engage with patient portals, our focus groups revealed the general acceptance of patient portals, which were described overwhelmingly as tools with the potential for providing exceptional, personalized care that may even work to mitigate the unfair burden of disease for Black Americans in primary care settings. Thus, opportunities for better health care will clearly arise with increased communication, experience, and adoption of remote health care practices among Black Americans.

(J Med Internet Res 2021;23(6):e27820) doi:10.2196/27820

KEYWORDS

health promotion; patient engagement; telehealth; telemedicine; health disparities; technology acceptance model; health belief model

Introduction

Background

The use of electronic health records (EHRs) is rapidly becoming the norm for primary care services. The American Recovery and Reinvestment Act of 2009 attempted to provide access to EHRs for every American within 5 years through US $19 billion in incentives in the Health Information Technology for Economic and Clinical Health Act [1]. According to the most recent National Electronic Health Records Survey from the Centers for Disease Control and Prevention in 2017, 85.9% of office-based doctors use EHRs—up by 37.6% from the year the Health Information Technology for Economic and Clinical Health Act was passed [2,3].

Patient portals, which were first developed by software vendors and health care centers in 1999, provide patients access to their medical charts and other data from their EHRs [4]. Patient portals vary by provider but many also facilitate payments, scheduling, and medication refills; link patients to educational materials; and allow password-secured communications with providers [5,6]. These tools are associated with increased effectiveness in the comprehensive care of chronic diseases, including diabetes and hypertension [7], which are both more prevalent and result in more severe outcomes among Black Americans compared with the rest of the American population. This health disparity is likely due to the psychosocial and socioeconomic stresses associated with the structural racism that is evident across the United States [8]. Proactive, portal-based secure messaging from the treatment team to the patient between visits has been shown to increase patient engagement and self-management behaviors in the treatment of chronic disease [9]. Despite the rapid adoption of patient portals at clinics and the positive outcomes associated with patient portal use, only 24.9% of the National Health Interview Survey respondents engaged with patient portal tools in 2017 [10]. Black Americans were less likely to activate their portal account and log-in and even less likely to use portal tools (12.2% of the Black National Health Interview Survey respondents) such as appointment scheduling, regardless of income or education [10,11]. Among Black Americans who are not registered for portal access, some barriers to portal use include a lack of support (ie, technological support and interaction with the treatment team); concerns for the loss of privacy, which are likely influenced by historical injustices within the health care system; and reduced personalization of care [12,13].

COVID-19 and the need to implement physical distancing to limit the spread of the associated virus have reduced the availability of in-person visits for primary and nonemergent care since early 2020 [12]. As of the second quarter of 2020, office-based visits were reduced by 50.2% [14]. In-person preventive care, chronic disease follow-up, and visits related to other primary care concerns can often be adapted to virtual care, although the frequency of routine assessments might change [14,15]. In place of in-person visits, primary care providers have increasingly used telemedicine and secure messaging to meet and communicate with their patients [12,14,15]. For example, one large primary care system asked 59% of its primary care staff to work from home by the end of March 2020, resulting in secure messaging to and from their 58,000 patients to increase by 41% and 51%, respectively [15]. In addition, telemedicine is also likely to increase the overall utility of patient portals because an increasing number of portals allow scheduling of telemedicine visits, access to the visit summary, and information on any recommended follow-up steps [5,6,15]. Following in the footsteps of federal agencies (ie, Medicare), private insurance companies have recently responded to the COVID-19 pandemic by lifting previous restrictions on telemedicine and by reducing or eliminating co-pays for telemedicine visits [12]. Although telemedicine is more accessible, the sudden shift to it has decreased the total number of appointments in primary care by 21.4%, according to the IQVIA (formerly Quintiles and IMS Health, Inc) National Disease and Therapeutic Index [14]. The frequency of assessments of cardiovascular risk factors and the prescription of new medications have also been reduced [14]. As Black Americans are already less likely to schedule office-based preventive care visits with their primary care provider [11,16], the need to digitize routine visits presents another barrier to scheduling, assessment, and treatment [14]. Digitization of health care will likely continue to grow even beyond the COVID-19 pandemic; thus, the importance of effectively connecting Black Americans to digital health care services is overwhelmingly necessary [17].

Objectives

We used focus groups to understand potential barriers to optimized care for conditions disproportionately affecting Black Americans, and participants repeatedly noted the impact of using portals on their care. The perceived utility and value of patient portals were often integral to participant experiences with care and their relationships with providers, health care institutions, and their own health, as detailed herein.

Methods

Study Design

Self-identified Black American residents aged 30-60 years with regular access to primary care were recruited nationally using Craigslist and Research Match advertisements. Recruitment continued until thematic saturation [18] was reached. Interviews were conducted remotely through Webex (Cisco Systems) in April and May 2020, with both the study team and participants located in their own environments. The participants were sorted into focus groups on a first-come-first-serve basis. Individual phone calls were implemented to answer any participant questions, collect demographic data, and obtain verbal informed consent per the Oregon Health & Science University Institutional Review Board exemption guidelines.

Each interview was an hour long, with a maximum of 5 participants and 2-3 facilitators. One facilitator was responsible for primary facilitation and asking probing questions and at least one facilitator took notes. The focus groups began with the facilitators describing their role and any personal connection to the research and reminding participants that the information disclosed in the interview would remain confidential. The participants then introduced themselves and provided a first name or nickname for use in the focus group to engage with other participants; no other personal identifiers were provided during the session. The subsequent discussion was participant led. At least one facilitator self-identified as Black in every session. Interviews were recorded with video and audio, transcribed verbatim, and deidentified by the primary facilitator.

Data Analysis

Patterns within the data were analyzed and identified through inductive thematic analysis using NVivo software (version 12.0; QSR International) [19]. An intercoder reliability analysis was performed using percent agreement among the raters [20]. Respondent validation was not used because of the quick evolution of the COVID-19 pandemic and the onset of nationwide protests that have involved many in the Black community and may therefore affect perspectives on health disparities or COVID-19 transmission and add additional participant burden [21]. The development of the interview guide was informed by grounded theory (Table 1). Study design and interview analysis were reported using the Consolidated Criteria for Reporting Qualitative Studies 32-item checklist (Table S1 in Multimedia Appendix 1) [22].

Table 1. Interview guide for 8 remote focus groups.

Prepared question	Interview numbers
Briefly describe a typical appointment with your primary care provider.	Interviews 1-3
Discuss the relationship between you and your primary care provider.	Interviews 1-8
When you have an appointment, how are recommendations or other health information conveyed to you?	Interviews 1-8
What are some health conditions that you think Black Americans are the most at risk of developing?	Interviews 1-3
Imagine that your primary care provider offered a new treatment or drug designed solely for Black Americans. How would you respond?	Interviews 1-3
If your primary care provider offered you a holistic plan to treat, for example, high blood pressure, including drugs, physical activity, and nutrition, would you be receptive to this? Why or why not?	Interviews 4-8
What are your concerns in the prevention and care of COVID-19?	Interview 4 and Interviews 6-8

Theoretical Framework

A total of two theoretical frameworks were integrated to direct the discussion of emergent themes related to patient portal use among the participants: the technology acceptance model (TAM) and the health belief model (HBM) [23]. The TAM posits that the perceived usefulness and ease of use of a technology determines attitudes toward and actual use of that technology [23]. The HBM is used to predict health behavior and posits that individuals with high perceived personal risk of disease are more likely to seek health information and engage in health behaviors [24]. The integration of these models may predict the likelihood of adopting portal technology (TAM) to support an individual's health awareness and provider communication and their engagement in healthy behaviors (HBM), including the adoption of preventive measures and treatment [25].

Data and Materials Availability

The data that support these findings are available on request from the corresponding author, NPB; the data are not publicly available because of ethical restrictions.

Results

Overview

Eight focus groups engaged 29 participants (22 women) with a mean age of 40 (SD 8) years. Intercoder reliability was found to be in ≥95% agreement, such that secondary codes were not used in the analysis. Of the 29 participants, 22 (76%) discussed experiences related to patient portals. From the inductive thematic analysis, four themes emerged regarding patient portal use among Black Americans in primary care: optimization of care, patient empowerment, patient-provider communication, and patient burden. Table 2 provides the frequency of these four themes, and in the *Optimization of Care*, *Patient Empowerment*, *Patient-Provider Communication*, and *Patient Burden* sections, we provide some of the most salient individual examples that emerged from discussion in each of these themes.

Table 2. Frequency of identified themes in 8 focus groups.

Theme	Frequency values[a]
Optimization of care	29
Empowerment	15
Patient-provider communication	34
Patient burden	9

[a] Frequency was calculated as the total number of thematic endorsements of a particular code by an interviewee in any number of interviews.

Optimization of Care

Most participants described patient portals as a tool to keep both the doctor and patient informed and organized, as well as to connect with the treatment team. For many, patient portals are integrated into the appointment routine and are considered a *go-to* tool for patients before and after each visit. Before the appointment, these portals were used for scheduling, requests for tests, insurance navigation, and communications. Scheduling is often the first point of contact for patients:

> *[The portal is] a really convenient tool especially for this day and age. I think that the doctor is more accessible that way. It's a lot harder to have to schedule an appointment and have to go [in-person]. And then if something gets in the way and you have to reschedule, it could all be delayed.* [Participant 4, group 1, Portland, Oregon]

After scheduling, the portal helped patients to prepare for their next visit. As noted by 1 participant, physicians may respond differently to requests for tests. For this participant, use of the portal before appointments complemented her own engagement with her health, as well as her new physician's willingness to work with her to understand her symptoms:

> *[The portal is] typically how I get my information...When I was desperately trying to find out what is wrong with me, I was always going online or finding something that I thought I might have. Interestingly enough, I would email my primary care physician via the web portal, and he would say "okay, well we'll test you for this or test you for that..." My last experience before I switched primary care physicians was most harrowing: I didn't feel heard, and I'm sure a lot of people have had this experience as well where they say "oh, you have this so, here, get this prescription," or "do this," or "do that." So, it's been quite a journey to finally find a diagnosis for what I have...I like [my new physician's] approach because he is kind of like "let's rule it out" versus, you know, "let me figure it out or let me tell you what I think you have...."* [Participant 23, group 4, Fontana, California]

Communication with the treatment team before the appointment reduced the possibilities of the participants forgetting or becoming too nervous to address specific health concerns:

> *I think you just have more time when you're on the computer to think of what you really [want to] tell the doctor....* [Participant 3, group 1, Denver, Colorado]

During appointments, the participants reported the efficiency of having personal and family health histories centralized and readily available to the treatment team through their EHR:

> *[The portal] also makes things quicker because [the doctor] could just be like "alright from your last visit I see this, and let's go ahead and move forward with that." I think because the doctor already knows what they're dealing with it probably makes it easier for them to make a more informed decision about [my] care... [When I] switch doctors it kind of feels like starting over...I think it helps having these uniform systems [and patient] information in the portal so [the new doctor] can at least catch up.* [Participant 4, group 1, Portland, Oregon]

After the appointment, treatment teams often contacted patients first (through phone, text, or secure email) for next steps. Participant access to portals for after-visit summaries and treatment recommendations (eg, prescription dosages and blood assay results) streamlined visit follow-up. Some platforms allowed participants to make payments and view insurance information:

> *[The portal is] my best good friend, medically speaking, because it's what I prefer to use to check my results. [My provider] also follows up with conversations via the portal, so I use the portal quite a bit...from payment options, to referrals, to the information she conveys to me—any recommendations she makes [when] we are no longer face to face, that's my primary go-to.* [Participant 13, group 3, Pasadena, Maryland]

Centralized patient health records, accessible to both the treatment team and the patient, limited redundancies and mistakes:

> *The portal is a very simple system, but I'm very thankful for that. [My primary care physician and I] can go back-and-forth and then if she tells me a certain dosage, I can just look on the portal—"oh what did she say?"* [Participant 15, group 6, Staten Island, New York]

For participants with relatively few health concerns, portal use reduced the number of appointments because test results and prescriptions are accessible on the web:

> *We use the portal a lot; sometimes I'm like three or four back-and-forth. In my mind this would've been a doctor's visit back in the day, but we go back-and-forth.* [Participant 15, group 6, Staten Island, New York]
>
> *I rarely have any other visits in-between. I like being able to go get my physical and then almost instantaneously get those results...delivered to me via the portal.* [Participant 17, group 5, Nashville, Tennessee]

The participants generally appreciated the integration of portals into their treatment:

> *I remember the time that you had to ask them for your results and stuff like that. I think it adds [to my treatment] because then you already have a sense of—sometimes before the visit—of things you're going to talk about. So, I think it's more helpful and empowering for patients to have that access.* [Participant 14, group 4, Portland, Oregon]

Patient Empowerment

Distrust was a major topic in the discussion of treatments for diseases that disproportionately affect Black Americans. Distrust was founded primarily on historical, personal, and media-based accounts of malpractice. Providers ignoring medication allergies or preferences was a commonly reported cause for distrust and switching providers. Under these circumstances, portals may facilitate the transfer of records. Patient portals also play a unique role in mitigating distrust, as participants described the importance of personal and family health histories—instead of, or in addition to, racial identity—in the consideration of a new treatment or therapy:

> *I've been with my primary care [provider] for years and years and years and I've been very vocal about what my concerns are, my family history, and my culture, and I'm already putting that stuff out there. [If] they come back to me like "oh, I agree with you, and per this research...," I would be more receptive, because it's like, "okay, you actually heard where I'm coming from. And you're listening to what my concerns are and why I have these concerns, and*

> *you're trying to work with me to create a solution." But if it's definitely from a place of privilege where it's like "well, you know, you Blacks, the fried chicken and you got—" it's like "oop, sorry, new doctor! That's what you're not going to do."* [Participant 23, group 4, La Mesa, California]

This experience highlights the desire for personalized interactions among the participants and the importance of providing educational resources that connect previous discussions between the patient and provider to any new treatment recommendations. Educational resources were often provided digitally after appointments through secure messaging or portal-based visit summaries. Others obtained this information on their own through internet searches.

User-friendly portals also empowered the participants to discuss sensitive matters more transparently and to engage in their own health care more effectively. Most portal users reported benefiting from additional control of their appointments, although some participants viewed the portals as a hurdle between themselves and their provider rather than an additional avenue for communication (see *Patient Burden* section). Thorough reporting of health concerns through the portal was especially useful before the visit because crossing the digital divide served as an icebreaker in what can sometimes be timid interactions with the treatment team:

> *[The portal] lends a platform for being more honest and I think people in general are less uninhibited online. So, I think if you're a patient you're probably going to be a bit more forthcoming about what your symptoms are, or what you're experiencing, or what's going on in order for the doctor to better be able to assess what's happening...I think we've all been in those situations at the doctor's office and it's awkward especially if it is a new doctor and you don't know them really well. So, for me, I feel the need to be more detailed as well, because it is through text so I'm trying to make sure they can follow.* [Participant 4, group 1, Portland, Oregon]

The participants also stressed the importance of documenting health concerns on the portals to reduce in-person anxieties that may arise when under time pressure:

> *I agree with what [Participant 4] said...I just have more time to think it out, and can write it down, and really get maybe even better feedback. Or, at least feedback you can look back on and tell somebody else if you want a second opinion.* [Participant 3, group 1, Denver, Colorado]
>
> *I didn't use [the portal] as much before because I thought it was kind of impersonal and dismissive... [Now I don't] have to store [health concerns] in my mind and wait for an answer, because maybe I was a little bit anxious about whatever was going on with me...it's helped take some of the anxiety out of me, especially as I started to develop a relationship with my new primary care physician and the nurse.* [Participant 23, group 4, Fontana, California]

Other anxieties potentially resulting in the participants forgetting or dismissing relevant points during appointments were commonly attributed to distrust based on experiences of being pressured by members of the treatment team or intimidation by a doctor's demeanor.

According to the participants, physician accountability reciprocated patient accountability with regard to portal use. Many participants reported experiences in which their primary care providers would give the impression, either subtly or explicitly, that they did not have time for concerns beyond those routinely addressed in annual visits. The participants often avoided provider dismissiveness when they expressed their concerns through the portal before the visit:

> *I like to have [my health concerns] in writing and I feel like it's going to hold the doctor more accountable having it in writing...the doctor gets back to you quicker [over the portals].* [Participant 3, group 1, Denver, Colorado]

One participant highlighted this theme in her experience of advocating for the health of her mother, an older adult. In this example, her mother's treatment team was described as neglecting the patient, but the information stored in the portal served as a reference for her family members. Using her proxy access to her mother's up-to-date portal, the daughter was able to contact the treatment team and make the necessary requests for tests that had previously been neglected:

> *When [my mom] was dealing with the MediCal, the people were overworked and underpaid...so I would go into my mother's portal and health insurance and I would look for the doctors that she needs and then I would basically spell out everything for them... I'd write down dates and everything that I did so that they can't turn around and say "oh, I forgot all this..." I definitely had to be the proactive one to go after them to show them if they were being not only insensitive but also they were giving information which was basically the opposite of what they said before. They were saying she has kidney issues and her hemoglobin [A1]C is elevated but then they didn't follow up with her for four months. And I would say, "how didn't you follow up when her diabetes is out of control?" But then you're saying "you have to play a role; you have to take this medication..." I was like "that's just not right. [There] was a huge void and a huge gap, and it worried me...It helps me understand more why anybody, but especially Black people are with the highest rates of diabetes, kidney disease, and amputations, and problems and not getting care...I can be here for [my mom], but a lot of people don't have that because they're working, or their kids don't live with them, or they [just] don't know."* [Participant 18, group 6, San Diego, California]

As portals provide a way to conquer the digital divide, physicians may also be encouraged to obtain second opinions and other professional insights more readily. Thus, patient portals may increase transparency for both patients and their providers:

> *I imagine that [nervousness] is similar for the doctor, where if they're in a face to face visit with you, it might be more awkward if they need to Google something or call a colleague. But if they're not face to face with you, it does give them a little more time and anonymity and in being able to go consult with somebody if they need to.* [Participant 4, group 1, Portland, Oregon]

Health history stored on the portal underscored the value of transparency for increasing care satisfaction and patient engagement. Physician access to health records through EHRs tailored primary care visits to the specific health history of the patient. The ease of access to health information, such as test results and family health histories, complemented participant efforts to maintain accountability over their personal health outcomes:

> *I have, you know, parents, family health issues, and I want to make sure I'm not impacted. So, I try to stay a little more on top of things...I am a little more diligent.* [Participant 15, group 6, Staten Island, New York]

Reported lifestyle changes were most often dietary or exercise-related. Commonly reported steps to counteract increased health risk were typically related to conditions that participants knew disproportionately affected Black Americans. Participants were specifically aware of racial disparities in the prevalence of diseases such as COVID-19, cardiovascular disease, and diabetes, and accountability for health was usually attributed to the participants' personal and family history of disease. Chronic disease was common among participants and their family members, and EHRs documented and facilitated discussions of these health concerns. The accessibility of health records made possible through most portals enabled a level of patient accountability necessary to address increased health risks:

> *I don't have to feel like I have to play politics, particularly with my PCP... [Regarding portal use]. Usually I initiate, but they're willing to give me the information that I need or want and allow me, without trouble, to make decisions. They don't treat me like oh you're the patient and you can't make the decisions...* [Participant 18, group 6, San Diego, California]

Some participants wanting a new provider felt too invested to find one if their family health records were stored locally because sharing this information with a new physician might be time consuming, as previously noted. Digital records enabled participants to change physicians without communicating family and personal health histories anew (if transferring to an interoperable site) and to obtain second opinions easily.

Patient-Provider Communication

Some participants expected patient portals to reduce personalization of care but instead found that the portals supplemented their interactions with their treatment team. The accessibility of portal communications enabled many participants to address their concerns more thoroughly and accurately, increasing their engagement with their treatment plans. Participants with negative experiences regarding the portal were often dissatisfied with their treatment team at the time and were sometimes seeking a new primary care provider. Overall, communication through portals mirrored communication patterns already established between the patient and their provider (eg, efficient and accommodating or terse and impersonal):

> *[The portal is] a double-edged sword, right?...The best way of thinking of it I think is that [if] I'm texting with my doctor, you can either give me a short answer or just say "yes/no" without any other follow up or results. And there's other instances that people use that as a "hey, I saw your message, I actually wanted to give you a call. After I get off the phone with you, I'm going to send you another message and we'll check in..." So, it all depends really more so on the doctor. If they use that as another touch point to stay in touch, that's good.* [Participant 23, group 4, La Mesa, California]
>
> *Assessments that were performed that day [are] delivered to me via [the] portal...I've also given them permission to notify me by email and by text....I'll get a [text] message, that might say "you have a new test result" and I like that, and I find it very helpful...I like the connectivity...Whenever I've had to reach out to my primary care physician, I get a pretty good return rate. You know, usually within 24-48 hours if I ever have any questions....* [Participant 17, group 5, Nashville, Tennessee]

Provider response time was generally important to participants, as was the thoroughness of response:

> *Pretty rapid response time but very minimal in the communication. Just very, like, terse, which is what it is. That was the only method of communication.* [Participant 16, group 6, Portland, Oregon]

Although communication through patient portals was common, portal use failed to supplement in-person communication for some participants. Ineffective communication patterns and delegation of follow-ups from the provider to other treatment team members (eg, medical assistant or nurse) compromised satisfaction with portal-based communications. Portal use hindered effective communication in at least two scenarios. First, when a phone call was still necessary to fill the request or answer a question:

> *It's a longer process for them to get into the portal when I can just call and say what I have to say, instead of waiting a week to get an answer.* [Participant 22, group 5, St. Louis, Missouri]

Second, when digital responses from the provider are delayed, insufficiently detailed, or confusing:

> *When the replies are delayed, that kills me the most. Like, "hey, I sent a message on Monday. Here it is, now Monday the following week. What's up?"* [Participant 23, group 4, La Mesa, California]

For many participants, the nurse or medical assistant was in charge of much of the digital communication with the patients. Delegation to other treatment team members limited digital communication with primary care providers for some participants compared with those who received secure emails and phone call check-ins from their provider. Although some participants were indifferent to delegated communications, personalized messages from the primary care provider were especially appreciated. Weekend, after-hours, and between-appointment communications represented a level of care that exceeded expectations.

Patient Burden

Portals provide a variety of tools but also rely on the patient to take advantage of these tools outside of the appointment. Some participants had trouble adapting. In addition to technological setbacks, some participants described the portals as an extra step to contact their treatment team, especially in cases where a phone call is inevitable. The sophistication of the portal interface may also play a role in burdening the patient because some participants reported that messages are not forwarded to their email, requiring them to log in proactively to check their inbox. A Portland, Oregon, participant recalled that she had experienced this issue with her past primary care provider:

> *Everything was communicated through a portal. But I had to actively go into the portal to see if there were communications. It wasn't like if she sent me a message and that would show up in my personal email and then I would know—I had to really be looking and waiting.* [Participant 16, group 6, Portland, Oregon]

Many participants were aware of the lack of interoperability among portals and were frustrated at having to recount family health history to new providers when switching insurance networks or moving:

> *I bring everything back to my primary care doctor. There's not a lot of communication between specialists. They do have their own little charts [records] and that can be hard.* [Participant 18, group 6, San Diego, California]

In addition to changing appointment routines to adapt to the portal, some participants with long-term relationships with their providers noted the burden that portals placed on hospitals and treatment team members to learn how to implement them in their practice. This physician burden resulted in at least 2 participants losing their long-term primary care provider when the portal was introduced at their place of care:

> *This doctor retired and then, I don't know, he said he was kind of almost forced out a few years back when everything was going digital and he didn't. He's on the older side so he didn't kinda learn the new system and stuff...I did have to switch to someone in the meantime, and I got a referral for her from a family member. But, yeah, she wasn't as accommodating I'd say, and that's why I went back to my first doctor.* [Participant 5, group 2, Portland, Oregon]

The participant chose the existing rapport she had with her very accommodating provider over a digitalized form of their relationship. Another participant noticed that she rarely received responses to secure messages and wondered if her treatment team had any training on the use of portal-based communications.

Discussion

Principal Findings

In the past, structural barriers have limited the initial portal sign-on by Black Americans [11-13]. In our focus groups with Black primary care patients who do use patient portals, the most frequently referenced themes were (1) optimization of care, (2) patient empowerment, (3) patient-provider communication, and (4) patient burden. The two most common themes related to patient portal use in our focus groups, namely optimization of care and patient empowerment, suggest that experiences with patient portals are largely positive. The influence of patient portal use on patient-provider communication and patient burden was generally contingent on previously established communication patterns, as well as support from the treatment team. Similarly, previously identified barriers to portal use among racial minorities who had little or no history of portal use include expectations regarding technological support and the quality of portal-based communications [13].

Technology Acceptance and HBMs

Applying the TAM, the emergent themes in our focus groups suggest that when the perceived usefulness and ease of use of patient portals improve, attitudes toward portals improve and so does the actual use of patient portals among Black Americans [23]. The perceived usefulness of patient portals for optimized care among our participants is further explained by the integration of the HBM with the TAM. According to the integrated model, individuals with high perceived susceptibility to disease and high health consciousness are more likely to adopt technology that fulfills health information and communication needs or that supports health-seeking behavior [24,25]. Meeting such needs in a way that is perceived as both easy and beneficial subsequently influences and reinforces positive attitudes toward and increased use of health technologies [23,25]. Susceptibility to disease was a common concern among the interview participants. In the face of historical mistreatment and ongoing bias against Black Americans in health care settings, portals may offer Black patients control over their appointments while providing tools for personalized care to mitigate bias. This finding is consistent with a March 2020 survey [26] showing that Black Americans who viewed COVID-19 as a minor personal health threat (as opposed to no threat at all to personal health or a major threat for which in-person assessments may be necessary) had the highest use of telemedicine as a result of the pandemic compared with White respondents, despite anticipated barriers. This may be due to increased perceived susceptibility to COVID-19 among Black Americans, resulting in an abundance of caution regarding in-person visits compared with White Americans with the same level of perceived threat [26].

Although telehealth may replace some elements of in-person visits, the results from this study suggest that portals may be relied upon for the optimization of care before, during, and after appointments when used by Black Americans. Once a telehealth or in-person appointment is made, the personalization of care may still be limited by implicit bias against Black Americans. Patient portals may help reduce this disparity by holding providers accountable for addressing specific health concerns documented on the web, rather than requiring patients to quickly recall and disclose sensitive information in person. Patients run the risk of making mistakes, feeling uncomfortable, and even being traumatized by having to repeat health information every time they meet with a provider when their personal health information is decentralized (spread across multiple local EHRs) or if the patient lacks an EHR altogether [27]. Black patients may feel this distress more poignantly [27], compounding existing distrust that is common among Black Americans in health care settings because of interpersonal and institutional racial discrimination [17]. Furthermore, centralized and up-to-date EHRs facilitate provider treatment recommendations based on personal and family health records and on environmental factors that have been documented, thereby improving health outcomes [1].

Looking Beyond Race

In contrast, race is often leveraged to assess risk in the treatment of Black Americans for conditions that they are at heightened risk for developing severe symptoms, including COVID-19 [28,29]. Race-based medicine and common risk assessments that use algorithms adjusted by a patient's race are present in cardiology, endocrinology, nephrology, urology, oncology, obstetrics, and other specialties [29]. This practice of operationalizing racial and ethnic categories results in the differential prescription of specialty services (including assessments, treatments, and major surgical procedures) for Black patients and other racial or ethnic minorities [29]. The prescriptive use of race in clinics may exacerbate inequalities and perpetuate implicit bias at health care institutions [17,29]. This contrasts with the use of race in descriptive statistics, which are vital for identifying disparities in health and for beginning to understand the etiology of disease (including systemic racism) and subsequent downstream effects (such as reduced access to care or increased psychosocial and environmental stressors) [8,29].

The impacts of race-based risk assessments for Black Americans at clinics include under- or overprescription of pain medications, reduced options for life-saving surgeries such as cesarean sections during childbirth because of preoperative risk assessments, less aggressive screenings for bone disorders and some forms of cancer because of lower risk of developing the condition (eg, breast cancer), and increased or decreased likelihood of intervention resulting from a decreased likelihood of survival (eg, rectal cancer treatment, in which doctors may not recommend treatment to patients who are unlikely to survive and recover) [29]. Although risk-assessment algorithms intend to increase efficiency in diagnosis and decrease costs, without proper scrutiny and understanding of the distinctions between biological mechanisms of disease and social determinants [8,17], adjustments by race might be arbitrary or even harmful or fatal [29]. In other words, racial identity is certainly not a replacement for a patient's health records. In our study, the focus group participants generally felt empowered by access to their own health information, including test results. Provider familiarity (or lack thereof) with family health histories was important to the participants, especially those considering looking for a new provider. This priority is consistent with the outcome of a community-based intervention for cancer-risk perception in Black Americans that emphasizes family health history to assess objective risk [30]. After the intervention, wherein family health histories were disseminated to the participants, objective and subjective risk levels matched in most of the participants [30]. Contextualizing subsequent treatments for the patient would likely require further collaboration with the patient to understand the barriers to effective care.

Portal-based communications emphasize the communication patterns already established between a patient and their provider. To this extent, portals provide the treatment team with additional opportunities for exceptional and tailored care. Care centered on the unique needs and experiences of the patient may be an expectation for many patients, particularly those with access to their own health records [1], but Black patients experience unfair bias in medical settings [17,29]. Predisposition to chronic disease and implicit bias in the treatment of such diseases place pressure on Black Americans seeking care. Indeed, recent work suggests that Black patients in primary care may feel that they must take great care to protect themselves from the health effects of structural racism while also confronting family history of disease through lifestyle choices and engaging more with their health [26,28]. By matching patient accountability with a level of care that exceeds expectations, regular portal use by patients and their providers can reduce bias and therefore alleviate some of the burden of disparity from the shoulders of Black patients [9]. Specifically, setting realistic and informed goals and then checking in proactively with the patient between appointments may address barriers in primary care with tools and information readily accessible on patient portals [6,9,31].

Strengths and Limitations

The potential study limitations include generalizability to populations lacking health coverage, access to internet-compatible devices, or regular access to patient portals. The themes discussed in this paper should be confirmed using quantitative methods, especially considering the increased use of telemedicine over the last year [12,14,15]. Web-based recruitment, as well as the web-based format of the interviews, also limits generalizability to individuals who are less computer savvy. The participants were aged between 30 and 60 years because of the requirements of the parent study of this investigation. Age is a limitation because older patients are likely to interact with their treatment team more often and may have more visits to manage. Despite the benefits of portals in managing care among providers, older populations may use portals less often than our participants and would require even more technological support and coordination between providers and other members of the treatment team [13,14]. As lack of technological support contributes to lower use of patient portals among Black Americans generally, our focus groups show how increasing support, communication, and adoption of patient portals may provide opportunities for better health care among Black patients.

Acknowledgments

This work was supported by a research supplement to R01 HL142064, the parent grant (SAS). Support was also provided by KL2TR002370 (NPB) and the Oregon Institute of Occupational Health Sciences at Oregon Health & Science University through funds from the Division of Consumer and Business Services of the State of Oregon (ORS 656.630). The parent study received support from UL1TR000128 (Oregon Clinical & Translational Research Institute). The authors would like to thank Shelby L Watkins (Masters in Public Health), for providing NVivo training to OHO and assistance with coding; Martina A Shannon for help with interview transcription; and members of the PreSERVE Coalition for Black/African American Memory & Brain Health who helped inform the interview questions and provided resources for parent study recruitment.

Authors' Contributions

RLC was involved in designing the interview guide and providing valuable mentorship to OHO throughout the study. LDR was involved in interview facilitation and probing. SAS is the principal investigator of the parent study and provided mentorship to NPB and OHO. NPB was involved in project design, interview facilitation, coding, and manuscript writing, as well as primary mentorship for OHO. NPB and OHO conceived the project, and OHO was involved in project design, primary interview facilitation, transcription, coding, and manuscript writing. All authors read, edited, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consolidated Criteria for Reporting Qualitative Studies 32-item checklist.
[PDF File (Adobe PDF File), 55 KB - jmir_v23i6e27820_app1.pdf]

References

1. Blumenthal D. Stimulating the adoption of health information technology. N Engl J Med 2009 Apr 09;360(15):1477-1479. [doi: 10.1056/nejmp0901592]
2. Hsiao C, Hing E, Socey T, Cai B. Electronic medical record/electronic health record systems of office-based physicians: United States, 2009 and preliminary 2010 state estimates. National Center for Health Statistics. 2010. URL: https://www.cdc.gov/nchs/data/hestat/emr_ehr_09/emr_ehr_09.pdf [accessed 2020-11-05]
3. National Electronic Health Records Survey. National Center for Health Statistics. 2017. URL: https://www.cdc.gov/nchs/data/nehrs/2017_NEHRS_Web_Table_EHR_State.pdf [accessed 2020-05-11]
4. Halamka JD, Mandl KD, Tang PC. Early experiences with personal health records. J Am Med Inform Assoc 2008 Jan 01;15(1):1-7. [doi: 10.1197/jamia.m2562]
5. Health IT: What is a patient portal? Office of the National Coordinator for Health Information Technology. 2017. URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2020-11-03]
6. Zarcadoolas C, Vaughon WL, Czaja SJ, Levy J, Rockoff ML. Consumers' perceptions of patient-accessible electronic medical records. J Med Internet Res 2013 Aug 26;15(8):e168 [FREE Full text] [doi: 10.2196/jmir.2507] [Medline: 23978618]
7. Zhou YY, Kanter MH, Wang JJ, Garrido T. Improved quality at Kaiser Permanente through e-mail between physicians and patients. Health Aff (Millwood) 2010 Jul;29(7):1370-1375. [doi: 10.1377/hlthaff.2010.0048] [Medline: 20606190]
8. Adler NE, Rehkopf DH. U.S. disparities in health: descriptions, causes, and mechanisms. Annu Rev Public Health 2008 Apr;29(1):235-252. [doi: 10.1146/annurev.publhealth.29.020907.090852] [Medline: 18031225]
9. Robinson SA, Zocchi MS, Netherton D, Ash A, Purington CM, Connolly SL, et al. Secure messaging, diabetes self-management, and the importance of patient autonomy: a mixed methods study. J Gen Intern Med 2020 Oct 21;35(10):2955-2962. [doi: 10.1007/s11606-020-05834-x] [Medline: 32440998]
10. Turner K, Hong YR, Yadav S, Huo J, Mainous AG. Patient portal utilization: before and after stage 2 electronic health record meaningful use. J Am Med Inform Assoc 2019 Oct 01;26(10):960-967 [FREE Full text] [doi: 10.1093/jamia/ocz030] [Medline: 30947331]
11. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 01;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
12. Das LT, Gonzalez CJ. Preparing telemedicine for the frontlines of healthcare equity. J Gen Intern Med 2020 Aug 3;35(8):2443-2444 [FREE Full text] [doi: 10.1007/s11606-020-05941-9] [Medline: 32495089]

13. 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. [doi: 10.2196/jmir.5910]
14. Alexander GC, Tajanlangit M, Heyward J, Mansour O, Qato DM, Stafford RS. Use and content of primary care office-based vs telemedicine care visits during the COVID-19 pandemic in the US. JAMA Netw Open 2020 Oct 01;3(10):e2021476 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21476] [Medline: 33006622]
15. Spelman JF, Brienza R, Walsh RF, Drost P, Schwartz AR, Kravetz JD, et al. A model for rapid transition to virtual care, VA connecticut primary care response to COVID-19. J Gen Intern Med 2020 Oct 23;35(10):3073-3076 [FREE Full text] [doi: 10.1007/s11606-020-06041-4] [Medline: 32705471]
16. Pullen E, Perry B, Oser C. African American women's preventative care usage: the role of social support and racial experiences and attitudes. Sociol Health Illn 2014 Sep 21;36(7):1037-1053 [FREE Full text] [doi: 10.1111/1467-9566.12141] [Medline: 24749849]
17. Tai D, Shah A, Doubeni C, Sia I, Wieland M. The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States. Clin Infect Dis 2021 Feb 16;72(4):703-706 [FREE Full text] [doi: 10.1093/cid/ciaa815] [Medline: 32562416]
18. Namey E, Guest G, McKenna K, Chen M. Evaluating bang for the buck. Am J Eval 2016 Jul 09;37(3):425-440. [doi: 10.1177/1098214016630406]
19. Alhojailan M. Thematic analysis: a critical review of its process and evaluation. West East J Soc Sci. 2012. URL: https://fac.ksu.edu.sa/sites/default/files/ta_thematic_analysis_dr_mohammed_alhojailan.pdf [accessed 2020-05-01]
20. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012:276-282. [doi: 10.11613/bm.2012.031]
21. Barbour R. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? Br Med J 2001 May 05;322(7294):1115-1117 [FREE Full text] [doi: 10.1136/bmj.322.7294.1115] [Medline: 11337448]
22. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
23. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
24. Champion V, Skinner C. The health belief model. In: Glanz K, Lewis FM, Rimer BK, editors. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, California, United States: Jossey-Bass; 1996:45-65.
25. Ahadzadeh AS, Sharif SP, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
26. 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]
27. Nordberg R. EHR in the perspective of security, integrity and ethics. Stud Health Technol Inform 2006;121:291-298. [Medline: 17095827]
28. Ordaz-Johnson OH, Croff RL, Robinson LD, Shea SA, Bowles NP. More than a statistic: a qualitative study of COVID-19 treatment and prevention optimization for Black Americans. J Gen Intern Med 2020 Dec 06;35(12):3750-3752 [FREE Full text] [doi: 10.1007/s11606-020-06259-2] [Medline: 33021713]
29. Vyas D, Eisenstein L, Jones D. Hidden in plain sight - reconsidering the use of race correction in clinical algorithms. N Engl J Med 2020 Aug 27;383(9):874-882. [doi: 10.1056/NEJMms2004740] [Medline: 32853499]
30. Murthy VS, Garza MA, Almario DA, Vogel KJ, Grubs RE, Gettig EA, et al. Using a family history intervention to improve cancer risk perception in a black community. J Genet Couns 2011 Dec 20;20(6):639-649 [FREE Full text] [doi: 10.1007/s10897-011-9389-2] [Medline: 21773879]
31. Boulware LE, Cooper LA, Ratner LE, LaVeist TA, Powe NR. Race and trust in the health care system. Public Health Rep 2003 Jul;118(4):358-365. [doi: 10.1016/s0033-3549(04)50262-5]

Abbreviations

EHR: electronic health record
HBM: health belief model
TAM: technology acceptance model

Edited by R Kukafka; submitted 08.02.21; peer-reviewed by C Jacob, A Hidki,Dr; comments to author 01.03.21; revised version received 17.03.21; accepted 06.05.21; published 03.06.21.

Please cite as:
Ordaz OH, Croff RL, Robinson LD, Shea SA, Bowles NP
Optimization of Primary Care Among Black Americans Using Patient Portals: Qualitative Study
J Med Internet Res 2021;23(6):e27820
URL: https://www.jmir.org/2021/6/e27820
doi:10.2196/27820
PMID:34081016

Original Paper

A Web-Based Time-Use Application to Assess Diet and Movement Behavior in Asian Schoolchildren: Development and Usability Study of My E-Diary for Activities and Lifestyle (MEDAL)

Airu Chia[1], BSc, PhD; Muhammad Naeem Jia Sheng Chew[1], BSc; Sarah Yi Xuan Tan[1], BSc; Mei Jun Chan[1], BSc, MPH; Marjorelee T Colega[2], BSc; Jia Ying Toh[2], BSc, MSc; Padmapriya Natarajan[1,3], MSc; Carla Lança[4], BSc, MSc, PhD; Lynette P Shek[5], MBBS, MRCP, MRCPCH, FAAAI; Seang-Mei Saw[1,4,6], MBBS, MPH, PhD, FAMS, FARVO; Falk Müller-Riemenschneider[1*], MBBS, MPH, PhD; Mary Foong-Fong Chong[1,2*], BSc, PhD

[1] Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore
[2] Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research, Singapore, Singapore
[3] Department of Obstetrics and Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[4] Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore
[5] Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[6] Duke-NUS Medical School, Singapore, Singapore
[*] these authors contributed equally

Corresponding Author:
Mary Foong-Fong Chong, BSc, PhD
Saw Swee Hock School of Public Health
National University of Singapore and National University Health System
12 Science Drive 2
Singapore, 117549
Singapore
Phone: 65 6516 4969
Email: mary_chong@nus.edu.sg

Abstract

Background: Web-based time-use diaries for schoolchildren are limited, and existing studies focus mostly on capturing physical activities and sedentary behaviors but less comprehensively on dietary behaviors.

Objective: This study aims to describe the development of My E-Diary for Activities and Lifestyle (MEDAL)—a self-administered, web-based time-use application to assess diet and movement behavior—and to evaluate its usability in schoolchildren in Singapore.

Methods: MEDAL was developed through formative research and an iterative user-centric design approach involving small groups of schoolchildren (ranging from n=5 to n=15, aged 7-13 years). To test the usability, children aged 10-11 years were recruited from 2 primary schools in Singapore to complete MEDAL for 2 weekdays and 2 weekend days and complete a 10-item usability questionnaire.

Results: The development process revealed that younger children (aged <9 years) were less able to complete MEDAL independently. Of the 204 participants (118/204, 57.8% boys, and 31/201, 15.4% overweight) in the usability study, 57.8% (118/204) completed 3 to 4 days of recording, whereas the rest recorded for 2 days or less. The median time taken to complete MEDAL was 14.2 minutes per day. The majority of participants agreed that instructions were clear (193/203, 95.1%), that MEDAL was easy to use (173/203, 85.2%), that they liked the application (172/202, 85.1%), and that they preferred recording their activities on the web than on paper (167/202, 82.7%). Among all the factors evaluated, recording for 4 days was the least satisfactory component reported. Compared with boys, girls reported better recall ability and agreed that the time spent on completing 1-day entry was appropriate.

Conclusions: MEDAL appears to be a feasible application to capture diet and movement behaviors in children aged 10-12 years, particularly in the Asian context. Some gender differences in usability performance were observed, but the majority of the participants had a positive experience using MEDAL. The validation of the data collected through the application is in progress.

(J Med Internet Res 2021;23(6):e25794) doi:10.2196/25794

KEYWORDS

time use; web-based; diet; movement behaviors; usability; schoolchildren

Introduction

Background

The prevalence of obesity in children and adolescents has increased in recent decades, and this rising trend has surged in parts of Asia [1]. In addition to short-term health consequences, overweight and obese children have a greater risk of adverse psychosocial health conditions and earlier onset of noncommunicable diseases [2]. Understanding factors contributing to obesity risk in children, such as physical inactivity and poor diet, is critical to prevent the accelerating burden of chronic diseases in later life.

However, there is a lack of good quality and timely data on schoolchildren, particularly those aged 10-14 years [3]. Parents may help to report the health behaviors of their children at home but often are not fully aware of their activities out of home, especially when children are increasingly independent in their food and activity choices at this age [4]. Although questionnaires completed by children are valuable tools to collect health behavioral data, they have the following limitations: (1) they often focus on a single health behavior; (2) they typically involve long-term recall, which is cognitively demanding for children (eg, examine activities during the past 7 days or last month) [5,6]; and (3) they rarely capture emerging health behaviors, such as nonscreen time sedentary behaviors, light-intensity physical activities, outdoor time, and tasks done concurrently (eg, screen time and eating) [7,8]. Although objective methods of measuring health behaviors are increasingly being used, they do not provide information on the type of activity or food consumed [9,10]. To address these limitations, a growing number of studies have used the time-use approach in children to collect information on health behaviors [11].

Time-use diaries prompt users to provide an account of a series of activities in a 24-hour day through previous day recalls [11]. Children are not cognitively developed to recall well before 8 years of age, but by the age of 10 years, they have the independent ability to provide accurate reports [12,13]. In addition, they allow the simultaneous capture of multiple energy balance–related behaviors. Most existing studies rely on paper-based diaries [11], whereas recent studies have developed software-based diaries [14-17], which have been demonstrated to engage children more effectively. However, to our knowledge, these studies focus mostly on capturing physical activities and sedentary behaviors but less comprehensively on dietary behaviors; for example, types of food consumed [14,17] and food portions [15,16] were not collected. Furthermore, there is a lack of application containing databases of Asian foods and common activities of children within the Asian context, such as tuition, also known as shadow education, which is prominent in East Asian countries [18]. In addition, the location of activities has been less studied, and no study to date has collected data on simultaneous activities (eg, eating while studying) in children aged 10-14 years, which is valuable for a better understanding of the activity contexts in children [19].

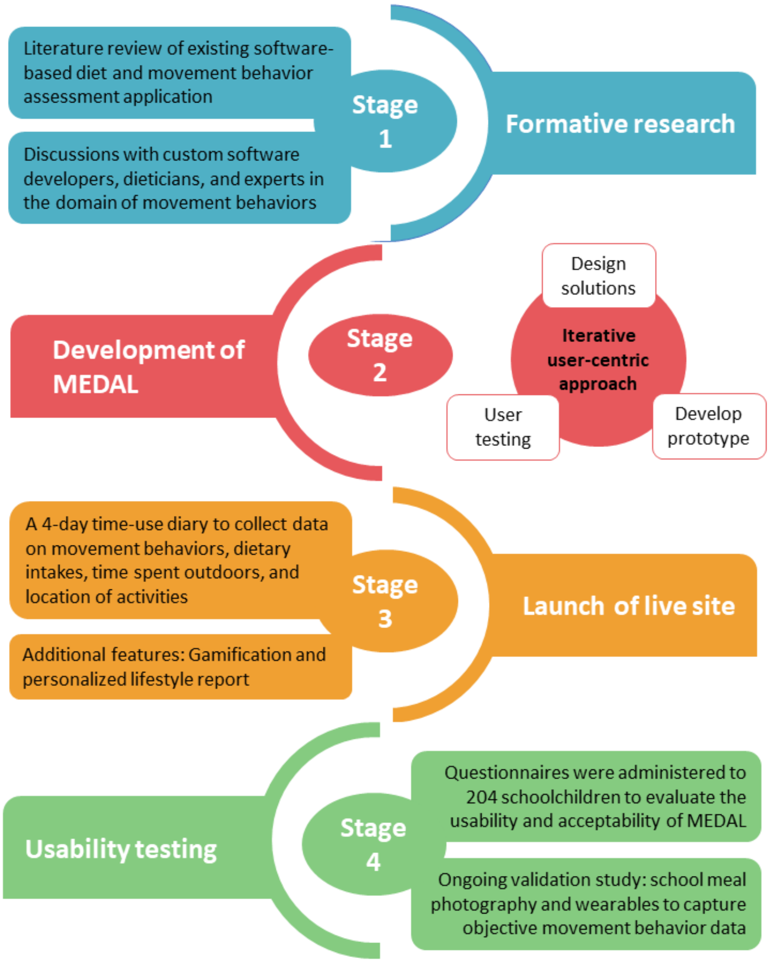
Objectives

To bridge these gaps, we developed a self-administered, web-based time-use application to assess diet and movement behaviors in schoolchildren: My E-Diary for Activities and Lifestyle (MEDAL). This study aims to describe the development of MEDAL and evaluate its usability among schoolchildren in Singapore.

Methods

The development of MEDAL was a multistage, user-centric process (Figure 1).

Figure 1. Flowchart illustrating the development process of My E-Diary for Activities and Lifestyle (MEDAL).



Stage 1: Formative Research

Several data sources were reviewed to identify the elements and requirements for developing a self-administered, web-based application for assessing diet and movement behaviors in schoolchildren. Both electronic and hand-searched strategies were adopted to identify existing software-based diet and physical activity assessment tools for schoolchildren. MyDailyMoves in the Netherlands [17], WebCAAFE in Brazil [16], Synchronised Nutrition and Activity Program in England [15], and Multimedia Activity Recall for Children and Adolescents in Australia [14] were evaluated. We also held interviews and meetings with experts from various knowledge domains, including custom software developers, experienced educators, and dieticians. The key points emanating were (1) application should be simple, easy to navigate, and take no more than 20 minutes per day to complete; (2) use of appealing visuals to engage children, and they prefer images to text and clicking instead of typing; and (3) enable children to report activities sequentially from wake-to-bed time (analogous to a time-use diary), so as to enhance recall—this method of *behavior chaining* has been shown to help children remember their activities [15]—and (4) complete the application at the end of each day to minimize omissions and intrusions [14,20]; children's accuracy in reporting has been shown to improve with shorter retention intervals (ie, time between event occurrence and reporting) [21].

Stage 2: Iterative User-Centric Development Process

On the basis of the formative research in stage 1, we adopted an iterative stepwise refinement approach to design, develop, and test features of the application [22]. User testing was conducted with children using methods such as paper-based simulations, audio clips, and mock-ups (ie, static wireframes) to understand the children's preferences for the interface and their comprehension of the task. At the end of user testing, user feedback was evaluated, and necessary improvements and amendments were conveyed to the software developers to inform the next development cycle. Multimedia Appendix 1 summarizes the key feedback and changes made during each development phase. Small groups of schoolchildren (aged 7-13 years), ranging from n=5 to n=15, were involved in all development stages to ensure that design decisions were tailored to their needs until the full functional system was launched. The sample sizes for each stage of user testing were adequate, as studies have shown that 80% of the usability problems were found with the first 5 users [23], and the diminishing return of uncovered problems was apparent when the sample size increased [24]. The key takeaways from the development phase were as follows: (1) recording should be limited to 4 days (similar to other diet and activity monitoring tools), as children found it tedious to record for 7 days; (2) prompts should be included to aid recall; (3) child-appealing visuals and intuitive designs should be used to ease navigation and motivate completion; and (4) younger children (<9 years) may need more assistance to complete MEDAL independently.

Stage 3: Launch of the Live Site

Technical Specifications

Multimedia Appendix 2 presents the major functionalities of MEDAL. MEDAL is a self-administered, web-based time-use application that takes users through the completion of a 4-day diary to collect data on movement behaviors, dietary intakes, time spent outdoors, and the location of these activities. It is written using the Yii Framework (version 1.0), PHP (version 5.6) as the application programming interface, Angular 2 (version 2) as the front end, MySQL (version 8) as the database, and hosted by the National University of Singapore. It is accessible on browsers such as Google Chrome, Microsoft Edge, Firefox, and Safari on devices such as tablets, laptops, and desktops.

Log-in and Registration

User-specific accounts are created by research administrators through the backend portal, and users can access MEDAL by logging on with a username and password given to them. Information on the date of birth and sex of the user is collected during the first log-in. MEDAL allows users to have the flexibility to record on any 4 days. Users can select if they are recording for the current day or the day before. The recall period is limited to the day before to minimize the retention interval and increase the likelihood of data validity [21].

Recording of Activities and Diet

Activities

Users are asked about the time they slept the night before and their wake-up time. Following this, users are asked to record their activities sequentially, with the end time of one activity leading to the start of the next, until the end of the day is reached (ie, bedtime). They can choose from 22 activities grouped into 6 broad categories: (1) shower or wash up, (2) eat and drink, (3) traveling, (4) sitting activities, (5) active activities, and (6) nap/sleep. For example, the traveling category included activities such as *walk*, *cycle*, *car/taxi*, and *bus/train*; sitting activities category included *reading/doing homework*, *hand phone/tablet*, *TV*, *video games*, *computer/laptop*, *board/card games*, *sit and chat*, and *tuition/music lesson*; and active activities included *run/jog*, *cycle/skate/scoot*, *ball games*, *martial arts*, *dance*, *neighborhood play, and water activities*. This list of activities was compiled from existing questionnaires examining the indoor and outdoor activities of schoolchildren [25]. Common activities listed in 7-day activity diaries completed by 200 children of a local multiethnic birth cohort study (Growing Up in Singapore Towards healthy Outcomes) [26] were also considered (Cheryl Zhang, MPH, unpublished data, May 2018). These activities were further grouped to reduce the number of response options (eg, basketball, handball, and football are collectively known as *ball games*), as a previous report showed that providing a large number of options for children was associated with poorer quality of data collected [27]. Children can specify activities that are not listed by typing in a free-text box labeled *other*.

When an *active activity* is selected, users will be questioned further about its intensity: “How tiring was the activity?” To aid children in conceptualizing activity intensity, 3 emoticons with captions illustrating different degrees of physical exertion are presented: (1) Just a little—you can sing and talk during the activity, (2) Quite tiring—you can talk but cannot sing during the activity, and (3) Very tiring—you cannot say more than a few words without pausing. Energy expended at each activity can be determined by multiplying the duration of the activity and its corresponding metabolic equivalent task, drawn from the Youth Compendium of Physical Activities [28].

Diet

When the category *eat and drink* is chosen, users will be asked to select the food and drink they consumed from 88 food items (icons) organized under 13 main food and drink groups (Multimedia Appendix 3). When necessary, items were placed in multiple food groups so as to enhance user accessibility (eg, French fries was located in the *noodle/pasta/potatoes* group and in the *fast food* group). Besides browsing the food and drink groups, users can perform a text search to display the food items as typed. The embedded fuzzy search feature allows items to be searched and displayed even when the food names are misspelled, thus greatly enhancing the ease of use of MEDAL for children. The list of food and drinks was compiled from commonly consumed foods identified from 3-day diet records of 360 local children of the Growing Up in Singapore Towards healthy Outcomes cohort [29] as well as discussions with experts in the field of dietary assessment. Users were also asked to choose the portion sizes of food that best corresponded to the amount they consumed from 4 pictorial options (shown simultaneously) per food item. A total of 4 pictorial options were provided, as this was found to be the optimal number of options for children to select from [20]. Moreover, the even-numbered options limit the middle option from being chosen out of convenience [30]. Simultaneous presentation of food images depicting increasing portion sizes is preferred over the sequential presentation for children to quantify their food intake more accurately [31] and within a shorter period [32]. For items that are not listed, users can type into the *other* textbox and indicate the amount consumed from the 4 options provided: half a portion, 1 portion, one-and-a-half portion, or 2 portions. Prompts for commonly forgotten foods “Did you forget to select any fruits, drinks, dessert, or supplements you have eaten?” were provided after each meal entry. Users also have the option to review and edit the food and drink items entered before recording the next activity.

Additional Features

MEDAL requests users to indicate activities that were performed simultaneously (eg, watching TV while having their meals), which may not be achievable in existing assessment tools [14-16]. Users were also prompted to enter the location of their reported activities (ie, home, school, student care, or others) and if they were done indoors or outdoors. Inquiry on location helps children to contextualize and recall the activities they did [15,20]. Following the selection of an activity, children were asked to record the end time of the activity. The recorded end time should be at least 5 minutes from the start time. The minimum time frame of 5 minutes was designed to allow children to record short bouts of activity [14]. To check for activities that may be missed out, prompts such as "That's an odd timing, are you sure you ended this at [time]?" will be provided if users recorded an activity that lasted for more than 3 hours. MEDAL automatically saves the user's responses and allows them to log in at any time of the day to continue recording or review and edit their activities if they are not able to complete in one sitting.

Gamification

Several game design elements were implemented to motivate children to complete MEDAL. Users can personalize the background interface and select their choice of avatar. The chosen avatar provides instructions and prompts in the form of visual texts and auditory dialogs, which serve to alert the user to action (ie, more reactive) and enhance user immersion in the MEDAL task [33].

Each completed day of recording rewards users with a gardening tool to nurture their virtual plant. At the end of 4 complete days of recording, a randomly assigned mystery fruit will be revealed to the user. These visual representations of achievements directly communicate the success or completion of the task and serve to evoke feelings of competence and efficiency in children [34]. The addition of the mystery element and randomization further aims to spark curiosity in children and motivate them to complete the task [35].

Personalized Lifestyle Report

On completing 4 days of recording, users will instantly receive a personalized lifestyle report. The report provides information on 4 components of movement behaviors—physical activity duration, outdoor duration, screen viewing duration, and sleeping duration—and evaluates the intake of 5 food groups—fruits, vegetables, whole grains, sugar-sweetened beverages, and dairy and dairy products. The mean durations or servings per day (ie, daily estimates) were calculated based on standardized algorithms and compared with the recommended guidelines [36-38]. Specific messages were delivered to users for each component based on the categories they belong: (1) met recommended guidelines, for example, "Terrific! Super-duper proud of you! You should continue to spend less time on the screen"; (2) nearly meeting guidelines, for example, "Almost there! Aim to spend less time on the screen. Keep going!"; or (3) did not meet guidelines, for example, "You should spend less time on the screen. Aim for less than 2 hours of screen time daily. Persevere! I believe in you!" Messages were gain-framed (ie, focused on attaining a desirable or positive outcome) and presented with graphical illustrations, as evidence from existing reviews and qualitative research has shown that this is effective in motivating children [39].

Stage 4: Usability Testing

Study Population

To test the usability of MEDAL, primary 5–level children (aged 10-11 years) from 2 coeducational government schools in Singapore were invited to participate in the usability study between April and August 2019. Children were eligible if they were able to understand and respond to the assessment in English. Of the 458 invited children from 13 classes (7 classes from school A and 6 classes from school B), 49.8% (228/458) were recruited. The main reasons for nonparticipation were *did not return consent form* (150/458, 32.8%), *parents did not consent* (46/458, 10%), *child was busy or not interested* (20/458, 4.4%), and *child was absent on the day of study* (14/458, 3.1%). The study was approved by the institutional review board of the National University of Singapore (Reference S-18-088). Approval was also obtained from the Ministry of Education, Singapore, to approach schools for participant recruitment (reference RQ17-18(02)B). Informed written consent was obtained from the parents or guardians, and all participants gave verbal assent.

Data Collection

Practical sessions were conducted for each class in the schools' computer laboratories by 2 trained researchers who demonstrated how to log in and navigate the MEDAL application. Participants started their first MEDAL entry using school computers and completed the rest of the day's entry at home. Participants were instructed to record their diet and activities over 2 specified weekdays and 2 weekend days. They were encouraged to complete the day's entry at the end of the day (ie, before bedtime) or the day after. The schools' computer laboratories were opened during recess and after school to allow participants with no access to the internet or computer at home to complete MEDAL. School meal photography during the recess period was also conducted, and devices to capture objective data on movement behaviors (ie, accelerometers) and outdoor time (ie, FitSight watch, which measures light intensity [40]) were provided for validation of data collected through MEDAL.

At the end of the study, participants were instructed to complete a 10-item usability questionnaire based on a 4-point Likert scale (strongly disagree, disagree, agree, and strongly disagree). The questionnaire was adapted from existing usability questionnaires [41,42] and kept short to avoid satisficing responses [43]. Questions were positively worded, direct, and specific to tailor to the children's cognitive and communicative capacities [27]. The key themes covered by the questionnaire included clarity of instructions, motivation to continue recording, ability to recall activities and food, recording duration, and overall experience of using MEDAL. Participants were also allowed to provide comments and suggestions at the end of the questionnaire.

Information on participants' sex and age was collected from MEDAL. The time taken for participants to complete MEDAL each day was calculated as the duration of the first and last entry input (ie, bedtime), summed and averaged over the number of recording days. Height and weight were measured biannually in schools, and the most recent records were obtained. Underweight (<5th percentile), normal weight (5th to 90th percentile), and overweight (>90th percentile) were classified according to the age- and sex-specific BMI reference of the local population [44]. Parents were asked to indicate their child's access to computers and the internet at home (yes or no) in the consent form.

Statistical Analyses

Participants' characteristics were summarized according to sex, and comparisons were assessed by the Fisher exact test (categorical) and Wilcoxon Mann-Whitney test (continuous). Differences in usability attributes between sexes were examined using the Fisher exact test. The usability questionnaire responses *agree* and *strongly agree* as well as *disagree* and *strongly disagree* were merged into 2 categories, *agree* and *disagree*, before analysis. The method has been shown to be more helpful in the description of Likert scale data [45]. All statistical analyses were performed using Stata 14 (StataCorp LP). Statistical significance was set at *P*<.05.

Results

Characteristics of Participants

Of the 228 participants, 204 logged on to MEDAL and completed the usability questionnaire and were included in the analysis (Figure 2). The cohort comprised 57.8% (118/204) boys, with a median age of 10.9 (25th percentile: 10.6 and 75th percentile: 11.2) years, and 15.4% (31/201) were overweight (Table 1). The median time taken to complete MEDAL was 14.2 minutes per day, and the proportion who completed 3-4 days of recording, 1-2 days of recording, and <1 day of recording were 57.8% (118/204), 20.1% (41/204), and 22.1% (45/204), respectively. Compared with boys, girls spent more time completing MEDAL (median 15 min per day vs 13 min per day; *P*=.03) and were more likely to complete 3-4 days of recording (*P*=.01).

Figure 2. Flowchart of participants. MEDAL: My E-Diary for Activities and Lifestyle.

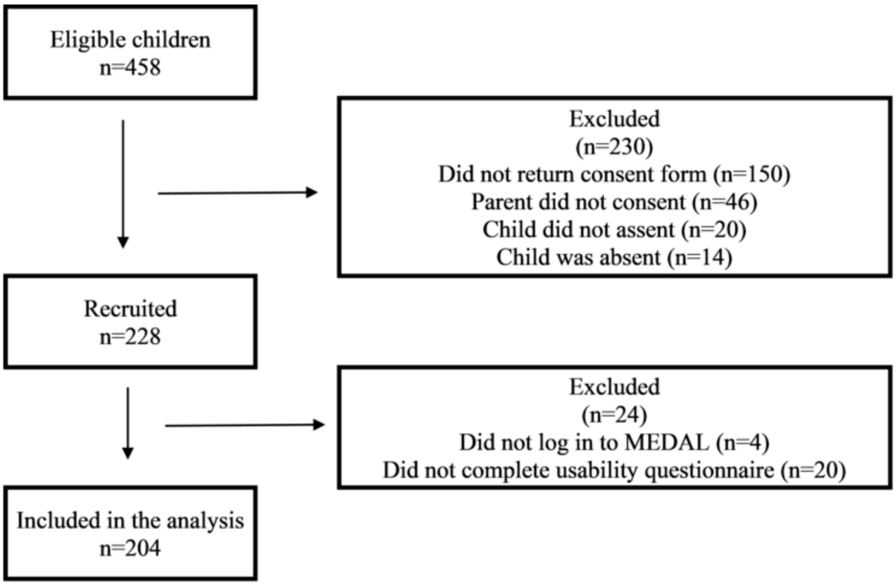


Table 1. Characteristics of participants summarized by sex (N=204).

Characteristics of participants	All (N=204)	Girls (n=86)	Boys (n=118)	*P* value[a]
School, n (%)				.07
School A	141 (69.1)	53 (61.6)	88 (74.6)	
School B	63 (30.9)	33 (38.4)	30 (25.4)	
Age (years), median (25th-75th percentile)	10.9 (10.6-11.2)	11.0 (10.8-11.2)	10.9 (10.6-11.1)	.03
Weight status, n (%)				.41
Underweight	25 (12.4)	12 (14)	13 (11.3)	
Normal weight	145 (72.1)	64 (74.4)	81 (70.4)	
Overweight	31 (15.4)	10 (11.6)	21 (18.3)	
Access to the internet and computer at home, n (%)				.99
Yes	170 (89.5)	73 (90.1)	97 (89)	
No	20 (10.5)	8 (9.9)	12 (11)	
Days completed, n (%)				.01
<1 day	45 (22.1)	12 (14)	33 (28)	
1-2 days	41 (20.1)	13 (15.1)	28 (23.7)	
3-4 days	118 (57.8)	61 (70.9)	57 (48.3)	
Average time taken to complete in minutes per day, median (25th-75th percentile)	14.2 (10.4-19.0)	15.0 (11.1-20.3)	13.1 (10.1-17.6)	.03

[a] *P* values were assessed by the Fisher exact test (categorical) and Wilcoxon Mann-Whitney test (continuous). There were missing data for weight status (n=3) and access to the computer and internet at home (n=14).

Usability of MEDAL

When assessing the usability of MEDAL, the majority (193/203, 95.1%) of participants agreed that instructions were clear, that MEDAL was easy to use (173/203, 85.2%), that they (172/202, 85.1%) liked the application, and that they preferred recording their activities on the web than on paper (167/202, 82.7%; Table 2). Recording for 4 days and ability to recall events were reported to be the least desirable among other factors (146/203, 71.9% and 154/203, 75.9%, respectively). Overall, the percentage agreement for the 10 questions ranged from 71.9% (146/203) to 95.1% (193/203).

Table 2. My E-Diary for Activities and Lifestyle (MEDAL) usability responses according to boys and girls.

Survey questions	Agree, n (%)[a]			*P* value[b]
	All (n=204)	Girls (n=86)	Boys (n=118)	
1. I understood the instructions for MEDAL[c] clearly	193 (95.1)	83 (96.5)	110 (94)	.52
2. It was easy to use and I do not require further instructions	173 (85.2)	76 (89.4)	97 (82.2)	.17
3. Earning tools to grow my plant makes me want to continue the game	159 (78.7)	75 (88.2)	84 (71.8)	.005
4. I was able to remember all the activities I did	154 (75.9)	70 (82.3)	84 (71.2)	.07
5. I was able to remember all the food and drinks I had	165 (81.2)	75 (87.2)	90 (76.9)	.07
6. The time I took to complete one day's entry was just right	149 (75.6)	71 (83.5)	78 (69.6)	.03
7. Recording for 4 days was just right	146 (71.9)	66 (77.7)	80 (67.8)	.15
8. I prefer recording my activities online than using paper	167 (82.7)	73 (85.9)	94 (80.3)	.35
9. I would play the game again	162 (79.8)	73 (85.9)	89 (75.4)	.08
10. Overall, I liked the game	172 (85.1)	76 (89.4)	96 (82.1)	.17

[a] Values shown are the sums of "Strongly agree" and "Agree."

[b] *P* values were assessed by Cochran-Mantel-Haenszel test. There were missing data for question 1 (n=1), question 2 (n=1), question 3 (n=2), question 4 (n=1), question 5 (n=1), question 6 (n=7), question 7 (n=1), question 8 (n=2), question 9 (n=1), question 10 (n=2).

[c] MEDAL: My E-Diary for Activities and Lifestyle.

Compared with boys, girls were more motivated by the gamification to complete recording on MEDAL (75/85, 88.2% vs 84/117, 71.8%; *P*=.005) and agreed that the time spent on completing 1-day entry was appropriate (71/85, 83.5% vs 78/112, 69.6%; *P*=.03; Table 2). Compared with boys, a higher proportion of girls reported being able to recall all the activities they did (84/118, 71.2% vs 70/85, 82.3%; *P*=.07) and food and drinks they consumed (90/117, 76.9% vs 75/86, 87.2%; *P*=.07).

Open-ended Suggestions

A total of 23 participants provided open-ended suggestions. Participants (n=8) who recorded MEDAL for ≤2 days preferred recording on paper rather than on the web (Textbox 1). It was also reported that there was insufficient time to complete MEDAL. For participants (n=15) who completed 3 or 4 days of recording, there were suggestions to include more food and activity options and to have the ability to edit timings of activities (n=3 girls). Participants also proposed to allow recording on MEDAL for more than 4 days (n=3 boys) and modifying the game design, such as inclusion of more avatars and games and wanting a reward that is different from a plant.

Textbox 1. Suggestions provided by 23 participants.

Participants Who Recorded for ≤2 Days (n=8)

- Prefer recording on paper rather than on the web
  - “We could use paper to do rather than online” [Girl]
  - “I do not want the activity online but in paper” [Boy]
  - “To use paper than online is easy. I don’t know” [Boy]
- Insufficient time to complete MEDAL
  - “This program is very interesting but a bit troublesome as I don’t have much time to finish the online diary” [Girl]
- Include more options
  - “More options. Like more activities.” [Boy]
  - “Study in school” [Boy]

Participants Who Recorded for 3-4 Days (n=15)

- Include more options
  - “Include more options for us to choose” [Girl]
  - “Have more food options” [Boy]
  - “Sometimes we just do nothing, so please add a button saying ‘nothing’” [Boy]
- Ability to edit timing of activity
  - “Delete stuff [cause (sic) I accidentally write what time I sleep wrongly but could not change]” [Girl]
  - “During the game, when putting in the timing slots, when I put some numbers in wrong it is impossible to change” [Girl]
  - “For example, if you forgot what you did but you suddenly remembered you can change what you wrote” [Girl]
- Allow recording on MEDAL for more than 4 days
  - “I would ask them if they wanted to do more” [Boy]
  - “Make the days longer” [Boy]
  - “The ability to do the entries for the days even further behind” [Boy]
- Modifying the game design
  - “Put more characters!” [Girl]
  - “I would change the reward given.” [Girl]
  - “...to do more and they could get another plant” [Boy]
  - “More games” [Boy]
- Allow recording of sleep past midnight
  - “Make time for 12 am when I sleep” [Boy]

Discussion

Principal Findings

In this study, we described the development and evaluated the usability of MEDAL—a 4-day web-based time-use application to assess diet and movement behavior in schoolchildren in Singapore. Approximately 60% (118/204) of the participants completed more than 3 days of recording, whereas the rest recorded for 2 days or less. The majority of the participants reported having a positive experience using MEDAL. Compared with boys, girls were reported to be more motivated by gamification, had better reported recall ability, and agreed that the time spent on completing 1-day entry was appropriate.

During the development process of MEDAL, we observed that the completion rate of MEDAL was significantly lower in younger children (aged 8-9 years) than in older children (aged 11-12 years). Younger children aged <9 years may need more assistance to complete MEDAL independently; thus, MEDAL is more suitable for children aged ≥10 years.

Response rates for multiday diaries have generally been an issue among children [46]. It is perceived as tedious to report for ≥2 days, especially when activity patterns were more or less similar on school days. Respondents may also report fewer activities on subsequent days as a means of decreasing the level of burden [47]. Indeed, in our study, we found similar issues with response rates, such that only 57.8% (118/204) of the participants completed 3-4 days of recording. Recording for 4 days was also the least satisfactory component reported in the usability assessment. To strike a balance between the reliability of data and minimizing respondent burden, the collection of 2 diary days for a large study (ie, 1 weekday and 1 weekend day) or 3 days for smaller-scale investigations (ie, 1 day from Monday to Thursday, Friday, and 1 weekend day) has been deemed adequate to obtain representative time-use data in children [14,46]. Large sample sizes may level out individual behavioral misclassifications that result from the short assessment period and would consequently give accurate population-level estimates [48]. For the next version of MEDAL, we plan to provide the flexibility to record between 1 and 7 days and optimize the application for mobile devices on both Android and iOS platforms to cater to the needs of different projects, and users can access the website on the go.

We observed that participants who recorded for <2 days were less likely to have access to a computer and the internet at home (*P*=.02, results not shown), and qualitative data showed that they preferred recording on paper rather than on the web. Participants (20/190,10.5%) with no access to the computer and internet at home may be less tech-savvy, thus exhibiting lower usability performance [16] and greater preference for print than digital. In the next version of MEDAL, we plan to integrate a frequently asked questions section to assist participants, especially those with lower computer knowledge and skills, to troubleshoot their problems and enable them to navigate the site with ease.

Existing studies have shown that recall ability is a common difficulty reported by children [14,16], although girls tend to outperform boys in episodic memory tasks (eg, everyday events) [49]. This is in line with our findings, where we observed trends toward greater reported recall ability in girls than in boys. Participants suggested including more food and drink and activity options and pictures (eg, household chores and shopping). This feedback will be incorporated into the next version of MEDAL, as images will be useful to serve as visual memory prompts to enhance recall in children.

Studies have shown that boys have lower attentional performance and poorer inhibitory control but perform faster than girls [50]. Consistent with these findings, we found that girls spent more time completing MEDAL (median 15 min per day vs 13 min per day), but they were more agreeable to the time spent on MEDAL compared with boys. Generally, the actual average time spent on MEDAL was comparable with similar questionnaires completed by children [15,16], which is sufficiently short to hold the users' attention [20,51].

With regard to gamification, existing studies did not include this element [14-17]. In our study, girls appeared to be more motivated by the virtual gardening game to complete MEDAL recording than boys. Some boys have suggested integrating more games and that they wish to do more so that they can receive more rewards. Introducing more games may distract users and steer them away from recording their daily activities attentively [52]. However, as most of the participants reported a positive evaluation for the in-built game, we plan to revise and accelerate the game progression rewards in the next version of MEDAL so that users are able to receive rewards more frequently. Users will be rewarded with gardening tools as they record their activities, and a mystery fruit will be revealed after each completed day of recording. On the basis of the feedback from the participants, we will also include more avatars and backgrounds for users to choose and personalize in the next version of MEDAL.

Strengths and Limitations

MEDAL is a novel and engaging application for schoolchildren aged ≥10 years. It includes a range of Asian food options that cater to the key ethnic groups in Asia and captures data on food portions, which allow us to have a detailed assessment of energy and nutrient intakes. This is useful, as evidence suggests that portion sizes among children are increasingly poorly regulated [53]. In addition to assessing dietary intake and movement behaviors, MEDAL also examines the location of activities and tasks performed concurrently to allow a better understanding of the context of the activities and enable interventions to be more targeted. The usability evaluation of MEDAL, including user testing during the development phase, involves several methodological strategies (eg, direct observation, interview, Likert scale questions, and open-ended feedback), which allowed us to obtain rigorous and detailed qualitative and quantitative data.

However, there are some limitations to this study. First, we did not have information on children's academic ability. However, our usability questionnaire showed that 95.1% (193/203) of the participants understood the instructions on MEDAL, which was not unexpected, considering that English is the main language of instruction in Singapore and primary education is compulsory.

However, we recognize that the cognitive abilities to recall events, capability to concentrate, and comprehension of questions may have an impact on usability testing [16] and acknowledge that children with poorer cognitive ability may experience some challenges in using MEDAL. Currently, users are only able to edit types of activities but not the timing of the activities. In addition, the current application prevents users from recording beyond 12 midnight. As late bedtimes have become pervasive in children [54], the ability to record activities beyond 12 midnight, along with the function to modify timing, will be made available in the next version of MEDAL. In addition, to cater to both observational and interventional study designs, we plan to incorporate the flexibility to hide the lifestyle report so that it will not induce behavioral change in participants of observational studies.

Conclusions

In conclusion, we developed MEDAL, which appears to be a feasible web-based time-use application for children aged 10-12 years and potentially useful for children above this age group. We observed some gender differences in usability performance, but the majority of the participants had a positive experience using MEDAL. The validity and reliability of MEDAL will be examined next, and the results of this study will be used to inform the next version of MEDAL to customize MEDAL for children in other parts of Asia, which includes translation to their native language and adaptation of application content to their local context.

Acknowledgments

The authors would like to thank the principals, teachers, and children from both primary schools for their participation and contribution to this study. The authors would also like to thank the Ministry of Education, Singapore, for their help in the recruitment of schools for this study. This study was supported by the National University Health System Summit Research Program Partnerships-Office of Deputy President Research and Technology; Saw Swee Hock School of Public Health; and Yong Loo Lin School of Medicine, National University of Singapore. Additional funding was provided by the corresponding authors' National University of Singapore start-up grant. The funders had no role in the design of the study; collection, analysis, and interpretation of data; or writing the manuscript.

Authors' Contributions

MFFC, FMR, SMS, and LPS conceived and designed the study. ARC and SYXT designed and developed MEDAL with inputs from PN, FMR (content related to movement behaviors), CL, SMS (content related to location), MJC, MTC, JYT, and MFFC (content related to diet). SYXT, MNJSC, and ARC collected the data with assistance from MJC, MTC, and JYT. MNJSC and ARC performed statistical analyses. ARC drafted the manuscript with inputs from MFFC, FMR, SMS, SYXT, JYT, PN, and CL. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Development process of My E-Diary for Activities and Lifestyle (MEDAL).
[DOC File , 54 KB - jmir_v23i6e25794_app1.doc]

Multimedia Appendix 2
Functionality of My E-Diary for Activities and Lifestyle (MEDAL).
[DOC File , 1121 KB - jmir_v23i6e25794_app2.doc]

Multimedia Appendix 3
Food items under the eat and drink category.
[DOC File , 68 KB - jmir_v23i6e25794_app3.doc]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 2017 Dec 16;390(10113):2627-2642 [FREE Full text] [Medline: 29029897]
2. Sahoo K, Sahoo K, Choudhury AK, Sofi NY, Kumar R, Bhadoria AS. Childhood obesity: causes and consequences. J Family Med Prim Care 2015;4(2):187-192 [FREE Full text] [doi: 10.4103/2249-4863.154628] [Medline: 25949965]
3. Azzopardi PS, Hearps SJ, Francis KL, Kennedy EC, Mokdad AH, Kassebaum NJ, et al. Progress in adolescent health and wellbeing: tracking 12 headline indicators for 195 countries and territories, 1990–2016. Lancet 2019 Mar;393(10176):1101-1108. [doi: 10.1016/s0140-6736(18)32427-9]

4. Poulain T, Vogel M, Meigen C, Spielau U, Hiemisch A, Kiess W. Parent-child agreement in different domains of child behavior and health. PLoS One 2020 Apr 9;15(4):e0231462 [FREE Full text] [doi: 10.1371/journal.pone.0231462] [Medline: 32271851]
5. Shim J, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. Epidemiol Health 2014 Jul 22;36:e2014009 [FREE Full text] [doi: 10.4178/epih/e2014009] [Medline: 25078382]
6. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003 Jun 01;37(3):197-206 [FREE Full text] [doi: 10.1136/bjsm.37.3.197] [Medline: 12782543]
7. Matthews C, Moore S, George S, Sampson J, Bowles H. Improving self-reports of active and sedentary behaviors in large epidemiologic studies. Exerc Sport Sci Rev 2012 Jul;40(3):118-126 [FREE Full text] [doi: 10.1097/JES.0b013e31825b34a0] [Medline: 22653275]
8. Tremblay M, Barnes J, González SA, Katzmarzyk P, Onywera V, Reilly J, Global Matrix 2.0 Research Team. Global Matrix 2.0: report card grades on the physical activity of children and youth comparing 38 countries. J Phys Act Health 2016 Nov;13(11 Suppl 2):343-366. [doi: 10.1123/jpah.2016-0594] [Medline: 27848745]
9. Satija A, Yu E, Willett WC, Hu FB. Objective measures are complementary to, rather than a replacement for, self-reported methods. Int J Obes (Lond) 2015 Jul 30;39(7):1179 [FREE Full text] [doi: 10.1038/ijo.2015.80] [Medline: 25924712]
10. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International Physical Activity Questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395. [doi: 10.1249/01.mss.0000078924.61453.fb]
11. Hunt E, McKay EA. What can be learned from adolescent time diary research. J Adolesc Health 2015 Mar;56(3):259-266. [doi: 10.1016/j.jadohealth.2014.11.007] [Medline: 25592884]
12. Riley AW. Evidence that school-age children can self-report on their health. Ambul Pediatr 2004 Jul;4(4):371-376. [doi: 10.1367/a03-178r.1]
13. Foster E, Adamson A. Challenges involved in measuring intake in early life: focus on methods. Proc Nutr Soc 2014 May;73(2):201-209. [doi: 10.1017/S0029665114000020] [Medline: 24555806]
14. Ridley K, Olds TS, Hill A. The Multimedia Activity Recall for Children and Adolescents (MARCA): development and evaluation. Int J Behav Nutr Phys Act 2006 May 26;3:10 [FREE Full text] [doi: 10.1186/1479-5868-3-10] [Medline: 16725055]
15. Moore HJ, Ells LJ, McLure SA, Crooks S, Cumbor D, Summerbell CD, et al. The development and evaluation of a novel computer program to assess previous-day dietary and physical activity behaviours in school children: The Synchronised Nutrition and Activity Program (SNAP). Br J Nutr 2008 Jun 01;99(6):1266-1274. [doi: 10.1017/s0007114507862428]
16. da Costa FF, Schmoelz CP, Davies VF, Di Pietro PF, Kupek E, de Assis MA. Assessment of diet and physical activity of brazilian schoolchildren: usability testing of a web-based questionnaire. JMIR Res Protoc 2013 Aug 19;2(2):e31 [FREE Full text] [doi: 10.2196/resprot.2646] [Medline: 23958804]
17. Hidding LM, Chinapaw MJ, Belmon LS, Altenburg TM. Co-creating a 24-hour movement behavior tool together with 9-12-year-old children using mixed-methods: MyDailyMoves. Int J Behav Nutr Phys Act 2020 May 14;17(1):63 [FREE Full text] [doi: 10.1186/s12966-020-00965-0] [Medline: 32410623]
18. Kim YC, Jung JH. Worldwide shadow education epidemic: from East Asia to western hemisphere. In: Shadow Education as Worldwide Curriculum Studies. Switzerland: Palgrave Macmillan; 2019:25-60.
19. Matthews CE, Berrigan D, Fischer B, Gomersall SR, Hillreiner A, Kim Y, et al. Use of previous-day recalls of physical activity and sedentary behavior in epidemiologic studies: results from four instruments. BMC Public Health 2019 Jun 03;19(Suppl 2):478 [FREE Full text] [doi: 10.1186/s12889-019-6763-8] [Medline: 31159761]
20. Biltoft-Jensen A, Trolle E, Christensen T, Islam N, Andersen L, Egenfeldt-Nielsen S, et al. WebDASC: a web-based dietary assessment software for 8-11-year-old Danish children. J Hum Nutr Diet 2014 Jan;27 Suppl 1:43-53. [doi: 10.1111/j.1365-277X.2012.01257.x] [Medline: 22594587]
21. Baxter SD, Guinn CH, Royer JA, Hardin JW, Mackelprang AJ, Smith AF. Accuracy of children's school-breakfast reports and school-lunch reports (in 24-h dietary recalls) differs by retention interval. Eur J Clin Nutr 2009 Dec 16;63(12):1394-1403 [FREE Full text] [doi: 10.1038/ejcn.2009.107] [Medline: 19756033]
22. Basil VR, Turner AJ. Iterative enhancement: a practical technique for software development. IIEEE Trans Software Eng 1975 Dec;SE-1(4):390-396. [doi: 10.1109/tse.1975.6312870]
23. Virzi RA. Streamlining the design process: running fewer subjects. Proceedings of the Human Factors Society Annual Meeting 2016 Aug 06;34(4):291-294. [doi: 10.1177/154193129003400411]
24. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. CHI '93: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems 1993 May:206-213. [doi: 10.1145/169059.169166]
25. Manios Y, Androutsos O, Moschonis G, Birbilis M, Maragkopoulou K, Giannopoulou A, et al. Criterion validity of the Physical Activity Questionnaire for Schoolchildren (PAQ-S) in assessing physical activity levels: the healthy growth study. J Sports Med Phys Fitness 2013 Oct;53(5):502-508. [Medline: 23903530]
26. Soh S, Tint MT, Gluckman PD, Godfrey KM, Rifkin-Graboi A, Chan YH, GUSTO Study Group. Cohort profile: Growing Up in Singapore Towards healthy Outcomes (GUSTO) birth cohort study. Int J Epidemiol 2014 Oct;43(5):1401-1409. [doi: 10.1093/ije/dyt125] [Medline: 23912809]

27. Bell A. Designing and testing questionnaires for children. J Res Nurs 2016 Aug 18;12(5):461-469. [doi: 10.1177/1744987107079616]
28. Butte N, Watson K, Ridley K, Zakeri I, McMurray R, Pfeiffer K, et al. A youth compendium of physical activities: activity codes and metabolic intensities. Med Sci Sports Exerc 2018 Feb;50(2):246-256 [FREE Full text] [doi: 10.1249/MSS.0000000000001430] [Medline: 28938248]
29. Sugianto R, Chan MJ, Wong SF, Shek LP, Tan KH, Chong YS, et al. Evaluation of a quantitative food frequency questionnaire for 5-year-old children in an Asian population. J Acad Nutr Diet 2020 Mar;120(3):437-444 [FREE Full text] [doi: 10.1016/j.jand.2019.09.021] [Medline: 31866358]
30. Lillegaard IT, Overby NC, Andersen LF. Can children and adolescents use photographs of food to estimate portion sizes? Eur J Clin Nutr 2005 Apr 9;59(4):611-617. [doi: 10.1038/sj.ejcn.1602119] [Medline: 15702127]
31. Subar AF, Crafts J, Zimmerman TP, Wilson M, Mittl B, Islam NG, et al. Assessment of the accuracy of portion size reports using computer-based food photographs aids in the development of an automated self-administered 24-hour recall. J Am Diet Assoc 2010 Jan;110(1):55-64 [FREE Full text] [doi: 10.1016/j.jada.2009.10.007] [Medline: 20102828]
32. Baranowski T, Islam N, Douglass D, Dadabhoy H, Beltran A, Baranowski J, et al. Food Intake Recording Software System, version 4 (FIRSSt4): a self-completed 24-h dietary recall for children. J Hum Nutr Diet 2014 Jan;27 Suppl 1:66-71 [FREE Full text] [doi: 10.1111/j.1365-277X.2012.01251.x] [Medline: 22616645]
33. Parker JR, Heerema J. Audio interaction in computer mediated games. Int J Comput Games Technol 2008;2008:1-8. [doi: 10.1155/2008/178923]
34. Sailer M, Hense JU, Mayr SK, Mandl H. How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. Comp Hum Beh 2017 Apr;69:371-380. [doi: 10.1016/j.chb.2016.12.033]
35. Kapp K. Gamification designs for instruction. Instructional-Design Theories and Models, Volume IV: Routledge; 2016:367-400.
36. Tremblay MS, Carson V, Chaput J. Introduction to the Canadian 24-hour movement guidelines for children and youth: an integration of physical activity, sedentary behaviour, and sleep. Appl Physiol Nutr Metab 2016 Jun;41(6 Suppl 3):iii-iiv. [doi: 10.1139/apnm-2016-0203] [Medline: 27306430]
37. Wong PT. Young children with myopia get it worse later in life. TODAY. 2016 Oct 06. URL: https://www.todayonline.com/daily-focus/health/young-children-myopia-get-it-worse-later-life-singapore-study [accessed 2021-05-24]
38. My healthy plate. Health Promotion Board Singapore. URL: https://www.healthhub.sg/programmes/55/my-healthy-plate [accessed 2021-05-24]
39. Williamson C, Baker G, Mutrie N, Niven A, Kelly P. Get the message? A scoping review of physical activity messaging. Int J Behav Nutr Phys Act 2020 Apr 15;17(1):51 [FREE Full text] [doi: 10.1186/s12966-020-00954-3] [Medline: 32295613]
40. Verkicharla PK, Ramamurthy D, Nguyen QD, Zhang X, Pu S, Malhotra R, et al. Development of the fitsight fitness tracker to increase time outdoors to prevent myopia. Transl Vis Sci Technol 2017 Jun 01;6(3):20 [FREE Full text] [doi: 10.1167/tvst.6.3.20] [Medline: 28660095]
41. Brooke J. SUS-A quick and dirty usability scale. In: Usability Evaluation In Industry. Reading, United Kingdom: Taylor & Francis; 1996:4-7.
42. Vereecken C, Covents M, Maes L, Moyson T. Formative evaluation of the dietary assessment component of Children's and Adolescents' Nutrition Assessment and Advice on the Web (CANAA-W). J Hum Nutr Diet 2014 Jan;27 Suppl 1:54-65. [doi: 10.1111/j.1365-277X.2012.01290.x] [Medline: 23980931]
43. Borgers N, de Leeuw E, Hox J. Children as respondents in survey research: cognitive development and response quality 1. Bull Sociol Method/Bulletin de Méthodologie Sociologique 2016 Jul 22;66(1):60-75. [doi: 10.1177/075910630006600106]
44. Body Mass Index (BMI)- for-age children and youth aged 6 to 18 years old. Health Promotion Board. URL: https://bukitbatoksec.moe.edu.sg/qql/slot/u537/Total%20Curriculum/Curricular%20Areas/PE/BMI_chart.pdf [accessed 2021-05-24]
45. Sullivan GM, Artino AR. Analyzing and interpreting data from likert-type scales. J Grad Med Educ 2013 Dec;5(4):541-542 [FREE Full text] [doi: 10.4300/JGME-5-4-18] [Medline: 24454995]
46. Ziviani J, Lim C, Jendra‐Smith D, Nolan D. Variability in daily time use: methodological considerations in the use of time diaries for children. J Occup Sci 2008 Jul;15(2):111-116. [doi: 10.1080/14427591.2008.9686617]
47. Stopher P, Greaves S, Clifford E. Multi-day Household Travel Surveys: Sampling Issues. 30th Australasian Transport Research Forum. 2007. URL: https://www.australasiantransportresearchforum.org.au/sites/default/files/2007_Clifford_Greaves_Stopher.pdf [accessed 2021-05-24]
48. van der Ploeg HP, Merom D, Chau JY, Bittman M, Trost SG, Bauman AE. Advances in population surveillance for physical activity and sedentary behavior: reliability and validity of time use surveys. Am J Epidemiol 2010 Nov 15;172(10):1199-1206. [doi: 10.1093/aje/kwq265] [Medline: 20855469]
49. Loprinzi P, Frith E. The role of sex in memory function: considerations and recommendations in the context of exercise. J Clin Med 2018 May 31;7(6):132 [FREE Full text] [doi: 10.3390/jcm7060132] [Medline: 29857518]
50. Tremolada M, Taverna L, Bonichini S. Which factors influence attentional functions? Attention assessed by KiTAP in 105 6-to-10-year-old children. Behav Sci (Basel) 2019 Jan 08;9(1):7 [FREE Full text] [doi: 10.3390/bs9010007] [Medline: 30626060]

51. Revilla M, Ochoa C. Ideal and maximum length for a web survey. Int J Mark Res 2017;59(5):557-565. [doi: 10.2501/IJMR-2017-039]
52. Blohm I, Leimeister JM. Gamification. Bus Inf Syst Eng 2013 Jun 14;5(4):275-278. [doi: 10.1007/s12599-013-0273-5]
53. Eck K, Delaney C, Leary M, Famodou O, Olfert M, Shelnutt K, et al. "My Tummy Tells Me" cognitions, barriers and supports of parents and school-age children for appropriate portion sizes. Nutrients 2018 Aug 08;10(8):1040 [FREE Full text] [doi: 10.3390/nu10081040] [Medline: 30096799]
54. Matricciani L, Olds T, Petkov J. In search of lost sleep: secular trends in the sleep time of school-aged children and adolescents. Sleep Med Rev 2012 Jun;16(3):203-211. [doi: 10.1016/j.smrv.2011.03.005] [Medline: 21612957]

Abbreviations

MEDAL: My E-Diary for Activities and Lifestyle

Edited by R Kukafka; submitted 17.11.20; peer-reviewed by M Chia, D Berrigan, D Dumuid; comments to author 03.01.21; revised version received 03.02.21; accepted 04.05.21; published 09.06.21.

Please cite as:
Chia A, Chew MNJS, Tan SYX, Chan MJ, T Colega M, Toh JY, Natarajan P, Lança C, Shek LP, Saw SM, Müller-Riemenschneider F, Chong MFF
A Web-Based Time-Use Application to Assess Diet and Movement Behavior in Asian Schoolchildren: Development and Usability Study of My E-Diary for Activities and Lifestyle (MEDAL)
J Med Internet Res 2021;23(6):e25794
URL: https://www.jmir.org/2021/6/e25794
doi:10.2196/25794
PMID:34106084

Original Paper

Using Virtual Patients to Explore the Clinical Reasoning Skills of Medical Students: Mixed Methods Study

Ruth Plackett[1], PhD; Angelos P Kassianos[1], PhD; Jessica Timmis[1], MBBS; Jessica Sheringham[1], PhD; Patricia Schartau[2], MBBS, PhD; Maria Kambouri[3], PhD

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

[2]Primary Care and Population Health Department, University College London, London, United Kingdom

[3]Institute of Education, University College London, London, United Kingdom

Corresponding Author:
Ruth Plackett, PhD
Department of Applied Health Research
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 20 7679 3291
Email: ruth.plackett.15@ucl.ac.uk

Abstract

Background: Improving clinical reasoning skills—the thought processes used by clinicians to formulate appropriate questions and diagnoses—is essential for reducing missed diagnostic opportunities. The electronic Clinical Reasoning Educational Simulation Tool (eCREST) was developed to improve the clinical reasoning of future physicians. A feasibility trial demonstrated acceptability and potential impacts; however, the processes by which students gathered data were unknown.

Objective: This study aims to identify the data gathering patterns of final year medical students while using eCREST and how eCREST influences the patterns.

Methods: A mixed methods design was used. A trial of eCREST across 3 UK medical schools (N=148) measured the potential effects of eCREST on data gathering. A qualitative think-aloud and semistructured interview study with 16 medical students from one medical school identified 3 data gathering strategies: Thorough, Focused, and Succinct. Some had no strategy. Reanalysis of the trial data identified the prevalence of data gathering patterns and compared patterns between the intervention and control groups. Patterns were identified based on 2 variables that were measured in a patient case 1 month after the intervention: the proportion of Essential information students identified and the proportion of irrelevant information gathered (Relevant). Those who scored in the top 3 quartiles for Essential but in the lowest quartile for Relevant displayed a Thorough pattern. Those who scored in the top 3 quartiles for Relevant but in the lowest quartile for Essential displayed a Succinct pattern. Those who scored in the top 3 quartiles on both variables displayed a Focused pattern. Those whose scores were in the lowest quartile on both variables displayed a Nonspecific pattern.

Results: The trial results indicated that students in the intervention group were more thorough than those in the control groups when gathering data. The qualitative data identified data gathering strategies and the mechanisms by which eCREST influenced data gathering. Students reported that eCREST promoted thoroughness by prompting them to continuously reflect and allowing them to practice managing uncertainty. However, some found eCREST to be less useful, and they randomly gathered information. Reanalysis of the trial data revealed that the intervention group was significantly more likely to display a Thorough data gathering pattern than controls (21/78, 27% vs 6/70, 9%) and less likely to display a Succinct pattern (13/78, 17% vs 20/70, 29%; χ^2_3=9.9; *P*=.02). Other patterns were similar across groups.

Conclusions: Qualitative data suggested that students applied a range of data gathering strategies while using eCREST and that eCREST encouraged thoroughness by continuously prompting the students to reflect and manage their uncertainty. Trial data suggested that eCREST led students to demonstrate more Thorough data gathering patterns. Virtual patients that encourage thoroughness could help future physicians avoid missed diagnostic opportunities and enhance the delivery of clinical reasoning teaching.

(J Med Internet Res 2021;23(6):e24723) doi:10.2196/24723

KEYWORDS

computer simulation; web-based patient simulation; computer-assisted instruction; educational technology; medical education; clinical decision support systems; clinical decision making; clinical reasoning; clinical skills; primary care; diagnosis

Introduction

Background

Clinical reasoning skills are defined as the thought processes used by clinicians to formulate appropriate questions and diagnoses and, therefore, are critical to providing quality health care [1,2]. Poor clinical reasoning skills have been associated with missed diagnostic opportunities and poor patient outcomes [3-6]. To address the need to improve clinical reasoning skills, there has been a call for more explicit teaching of clinical reasoning skills in undergraduate medical education [7,8]. However, the optimal method of teaching clinical reasoning skills is not well understood because of the complexity of the skills and how they vary depending on context, knowledge, and experience [1,2]. Traditional methods of teaching clinical reasoning skills, such as clinical placements, place a considerable burden on faculty time and resources. Furthermore, growing numbers of students can result in fewer opportunities for exposure to a variety of clinical cases [9-11].

The use of digital teaching methods has been recommended to address gaps in clinical reasoning skills teaching and complement traditional face-to-face methods [8,12-15]. Virtual patients, a specific type of computer program that simulates clinical scenarios, has been recommended as an effective method [9,16,17]. Virtual patients allow students to be exposed to a large number of varied patient cases, which can help them develop their knowledge and create more complex mental representations of illnesses [18,19]. Learning through experience, reflection, and deliberate practice can also help students to develop and retain their skills [12,20,21]. Virtual patients are also becoming increasingly similar to clinical practice, as more consultations are being undertaken on the internet [22-25].

The Electronic Clinical Reasoning Educational Simulation Tool

To address the need for more structured clinical reasoning training using digital methods, electronic Clinical Reasoning Educational Simulation Tool (eCREST) was developed by the authors and web designers, Silver District [26], and is reported in detail elsewhere [27]. eCREST sought to influence 3 cognitive biases that have been found to influence clinical reasoning: the unpacking principle, confirmation bias, and anchoring [28-30]. The unpacking principle is the tendency to not elicit the necessary information to make an informed diagnosis. Confirmation bias is the tendency to seek information only to confirm a diagnosis. Anchoring is the tendency to stick to an initial diagnosis despite contradictory information [31]. To address these biases, eCREST was primarily focused on improving data gathering skills and flexibility in thinking about diagnoses rather than all clinical reasoning skills. In eCREST, students were presented with 3 videos of virtual patient cases. These patients presented to primary care with nonspecific respiratory symptoms, such as cough, which could be indicative of serious conditions such as lung cancer. The students were required to ask the patient questions from a list, received a video response, and formulated diagnoses and a management plan. To address potential biases, reflection was prompted at regular intervals throughout the case by asking students to revise their diagnoses, and they received feedback at the end of each case [32].

A trial evaluated eCREST in 3 UK medical schools to test feasibility and acceptability and is described in detail elsewhere [33]. This trial found that eCREST appeared to influence students' data gathering but had less impact on flexibility in thinking about diagnoses. Students in the intervention group appeared to show a more thorough data gathering pattern than controls, as they did not miss important information, but there was suggestive evidence (not statistically significant at 5%) that they may ask more irrelevant questions than controls. However, the quantitative data from the trial provided little further insight into how students gathered information while using eCREST and other data gathering patterns. In addition, little is known about how students gather information from previous clinical reasoning studies, as paper vignettes were used that did not require students to gather information [34,35]. A greater understanding of how students gather data when using virtual patients in real time will inform educators about how they can best support students in developing these skills. It could also help developers to design virtual patients that provide better training on data gathering skills. Therefore, this study aims to:

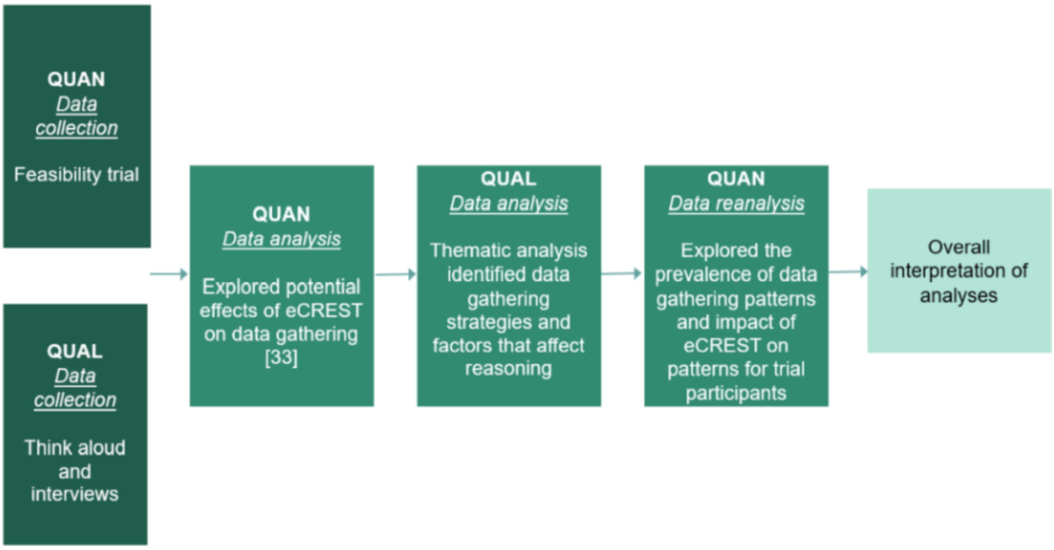
1. Understand how medical students gather information and reach diagnostic judgments when interacting with virtual patients.
2. Identify students' data gathering patterns while using virtual patients.
3. Examine whether eCREST changes the data gathering patterns of students.

Methods

Design

This study used a mixed methods design, as shown in Figure 1 [36-38]. The quantitative method was of equal priority to the qualitative method, and data collection was carried out concurrently and analyzed sequentially. Initially, a trial of eCREST was conducted to explore its potential effects on data gathering. The results of the trial's feasibility and effects on data gathering are reported elsewhere [33]. The methodological details of this study are also summarized in the sections below. During the trial, think-aloud protocols captured students' reasoning during eCREST and their reflections on the task. Following the initial analysis of trial data, qualitative data were thematically analyzed and distinct planned strategies of gathering data and factors that affect these strategies were identified (aim 1). This led to a reanalysis of the trial data to identify the prevalence of different data gathering patterns or behaviors among trial participants (aim 2). Finally, an exploratory analysis was undertaken to examine the effects of eCREST on data gathering patterns (aim 3).

Figure 1. Process of the mixed methods approach. QUAL: qualitative; QUAN: quantitative.



Setting

The trial took place at 3 UK medical schools, with full details reported elsewhere [33]. The concurrent qualitative study took place at one of the UK medical schools participating in the trial, with students who were not part of the trial. Ethical permission for the qualitative study was granted by the University College London (Ref. 9605/001; September 08, 2017).

Participants

For the trial, 18.16% (264/1454 of all eligible) final year medical students were recruited and randomly assigned to an intervention or control group. A total of 148 students (78 in the intervention group and 70 in the control group) remained in the trial after 1 month [33]. For the qualitative study, 16 final year medical students were recruited through a peer-assisted learning scheme, where the students opted to undertake a project in medical education, via newsletters and snowball sampling. All students were informed that their data would be anonymized and that their participation was voluntary, and they provided written consent. They were incentivized to participate with a voucher.

Procedure

Quantitative Data Collection and Initial Analysis

In the trial, the intervention group received 3 patient cases via eCREST that they had 1 week to complete while receiving teaching as usual. The control group received teaching as usual without access to eCREST until the end of the trial. Clinical reasoning skills outcomes were compared between the groups after 1 month through an additional eCREST patient case. Multimedia Appendix 1 shows a flowchart of the trial. The results indicated that eCREST may encourage more thorough data gathering patterns [33].

Qualitative Data Collection and Analysis

We used a think-aloud study design followed by semistructured interviews [39,40]. The think-aloud tasks involved observing students verbalizing their thoughts in real time while completing one patient case in eCREST. This method can provide insights into the clinical reasoning of medical students, as it provides access to their conscious thought processes [13,39,41,42]. A practice think-aloud task was given before the main task to ensure that students were comfortable with the process. Nondirective prompts, such as *keep talking*, were used if students were silent for a significant amount of time. Semistructured interviews were conducted after the think-aloud task to gather students’ retrospective thoughts [39,40]. A topic guide was developed for the interview and piloted with 2 medical students. Each session took approximately 1.5 hours per student.

All 16 think-aloud tasks and semistructured interviews were transcribed and imported into NVivo Version 12 (QSR International) software [43]. We used thematic analysis following the approach of Braun et al [44] and extended upon by Swain [45]. First, we developed deductive codes based on our research aims and initial findings from the trial data. We then familiarized ourselves with the data and developed inductive codes. A codebook was generated by one researcher (RP). To ensure the validity of the coding framework, 3 additional researchers (MK, APK, and JT) used the coding framework to guide their coding of a transcript and generated their own additional codes. We grouped similar codes into themes. Themes were validated in meetings with a wider research team where discrepancies were discussed, and a consensus was reached. Data were analyzed iteratively until the themes reached saturation. The results relating to how students gathered data (aim 1) informed the identification of students’ data gathering patterns in the larger quantitative data set obtained from the trial.

Reanalysis of Quantitative Data

Informed by the qualitative findings, we used data from the trial to estimate the prevalence of 3 data gathering patterns (aim 2) and to understand if eCREST may have changed these patterns (aim 3). Student data on an eCREST patient case completed by both interventions and controls 1 month after registration were used to assess clinical reasoning. As described in the trial paper, we constructed variables from these data to capture elements of clinical reasoning skills, such as data gathering ability and flexibility in thinking [33]. In this study, we focused only on the variables related to data gathering skills, as these were found to significantly differ between groups in the trial and the focus of this research was on data gathering skills. We assessed *essential information identified* (Essential) by measuring the proportion of essential questions they asked the patient and patient examinations undertaken, out of all possible essential examinations and questions identified by experts. This measure captured how complete data gathering was irrespective of whether irrelevant questions were also asked. The *relevance of history taking* (Relevant) was measured by assessing the proportion of all relevant and essential questions they asked the patient and patient examinations undertaken, out of the total number of examinations undertaken and questions asked. This measure was an indicator of the specificity of information gathered. Students could select a total of 70 questions and examinations. Experts defined 29 Essential items as a question or an examination that would change the differential diagnosis of the patient case, allow for differentiation as much as possible between alternative diagnoses, and reveal the key symptoms that might be indicative of sinister diagnoses. A further 10 questions were considered by experts as *relevant*, that is, clinically appropriate to ask but would not reveal key information to derive all important possible diagnoses. The remaining 31 questions were defined as *irrelevant*.

To identify the data gathering patterns displayed by students, the quartiles for the Essential and Relevant variables were calculated. Those who scored in the top 3 quartiles on both variables were classified as having a Focused pattern. Those who scored in the top 3 quartiles for Essential but in the lowest quartile for Relevant were classified as having a Thorough pattern. Those who scored in the top 3 quartiles for Relevant but in the lowest quartile for Essential were classified as having a Succinct pattern. Those whose scores were in the lowest quartile on both variables were labeled Nonspecific. Sensitivity analyses showed that other cutoff values below the lowest quartile had insufficient numbers for each pattern to conduct the chi-square analysis and compare scores between the intervention and control groups. We examined whether the use of different data gathering patterns varied between the intervention and control groups using a chi-square test. Analyses were conducted using STATA Version 15 with $P \leq .05$, considered statistically significant [46].

Results

Sample Characteristics of Qualitative and Quantitative Studies

Table 1 describes the participant characteristics of those in the trial at baseline and the think-aloud study. In the trial, most participants were 23 to 24 years of age (152/264, 57.6%) and were male (142/264, 53.8%). The age and gender of the think-aloud participants were similar: 81% (13/16) were 23 to 24 years old and 56% (9/16) were male.

Table 1. Participant characteristics at baseline in the trial and think-aloud study.

Characteristics	Trial			Think-aloud study
	Intervention group (n=137), n (%)	Control group (n=127), n (%)	*P* value[a]	Whole group (n=16), n (%)
Age (years)				
20-22	4 (2.9)	1 (0.8)	N/A[b]	0 (0)
23-24	73 (53.3)	79 (62.2)	N/A	13 (81)
25-26	39 (28.5)	29 (22.8)	N/A	2 (13)
27-28	11 (8)	10 (7.9)	N/A	1 (6)
>29	10 (7.3)	8 (6.3)	.49	0 (0)
Gender				
Female	64 (46.7)	58 (45.7)	N/A	7 (44)
Male	73 (53.3)	69 (54.3)	.87	9 (56)

[a] Comparisons between the intervention and control groups for the trial were made using chi-square tests. $P<.05$ was considered significant.

[b] N/A: not applicable.

How Students Gathered Information and Reached Diagnoses

From the qualitative think-aloud data, 4 major themes were identified relating to how students gathered information: data gathering strategies, structure of eCREST, diagnostic hypotheses, and confidence and uncertainty.

Theme 1: Data Gathering Strategies

Students had different data gathering goals and strategies for gathering information.

Being Thorough

Some students reported that they aimed to be thorough when gathering data and were aware of the potential pitfalls of the unpacking principle and not gathering all necessary information. Consequently, they asked many questions to reassure themselves that they had not missed any relevant information and symptoms indicative of serious disease. However, they acknowledged that this approach to data gathering could be lengthy and possibly led to them asking irrelevant questions:

> *I should probably go through those but I don't know, I do feel like I like to be thorough and I do want to ask all of the questions.* [P4]
>
> *I think I could be more concise. Cos, I just kind of ask everything just in case.* [P16]

Being Focused

Other students reported that they aimed to be focused when investigating the patient's symptoms and wanted to ensure that all the information they asked for was relevant to the patient case:

> *I don't really want to ask any more of these I mean partly because I feel like...I want to be focused, so I'm not really going to ask the rest of these which are potentially not that related.* [P10]

Being Succinct

Some also described that they aimed to be succinct and limit the number of questions they asked the patient, possibly because the way eCREST was structured or perhaps to be more time efficient. This shows that some students were less aware of the potential negative consequences of the unpacking principle and the importance of gathering all relevant information:

> *Maybe I should have not tried to limit myself to a specific number and asked what I thought was actually appropriate.* [P2]

Random Selection

Some students reported that they were asking questions randomly and had no discernable reasoning behind the way they gathered information. Students may have adopted this seemingly random approach because they were less engaged with patient cases in eCREST:

> *So these questions I guess aren't like very helpful but because they're there I'm going to ask them anyway.* [P11]

Theme 2: Structure of eCREST

The way in which eCREST was structured appeared to be one mechanism by which eCREST influenced how students gathered information. Students perceived eCREST to have both positive and negative effects on how they gathered information.

Organized Data Gathering

Students reported that the structure of eCREST helped them organize the way they gathered information by asking students to chunk their history taking into sections and regularly visualizing their diagnostic hypotheses. This helped them to hone their questions toward their diagnoses and may have helped some students to overcome the unpacking principle to take a more focused approach:

> *I guess it's good because it makes you streamline your thoughts regarding diagnosis after you have limited information available and I think that probably helps time management within GP settings, because it's making you streamline your questions.* [P9]
>
> *I think it was useful even for me to just like, to see when I'm taking the history, I think that...I don't really write down...the top five differential diagnosis when I'm taking history...Yeah, just better at visualizing it, and organizing it.* [P13]

Some appeared more focused in their approach to gathering data, as they thought of their own relevant questions first and then used eCREST as a checklist to confirm that they had not missed anything:

> *It's probably useful to try and think about this before I look at the list.* [P1]

Unrealistic Data Gathering

However, others reported that the list of questions in eCREST led them to ask questions more randomly and less strategically than they would in a real consultation:

> *I think it just biased the way in which I asked the questions. Because I wouldn't have just kind of gone through it and clicked on it as I went through.* [P2]

Some students felt that the lack of open questions in eCREST hindered their ability to gather information and felt open questions would have offered more relevant information in real life:

> *It's quite useful to make you think about the questions but because I don't ask questions in that, I feel like my own style is quite different to the way it's set out here...I'd be quite like open with the patient. I'd be like "tell me more." And then I'd be able, I'd have some better idea, I'd have a better timeline of the things.* [P15]

Theme 3: Diagnostic Hypotheses

The way in which eCREST guided how students generated and reassessed diagnostic hypotheses appeared to be another mechanism by which eCREST influenced how students gathered information.

Early Generation of Diagnostic Hypotheses

Some students found that eCREST helped them to think of diagnoses at an earlier stage than they usually would, which potentially helped them to avoid anchoring on a particular diagnosis early in the consultation and had an impact on the questions they asked:

> *Usually I don't really think about differentials so early on in a consultation...so this encouraged me to rule out different diagnoses from a very early point.* [P12]

Cognitive Biases

Some students became fixated on one or two initial diagnoses (anchoring) and would consequently seek information to confirm the diagnosis and stopped investigating other causes of the symptoms (confirmation bias). This may explain why some students took a more succinct approach to gathering information:

> *[Question: what will you try to improve?] Consider everything the patient has said and I think just not try to make diagnosis fit, like the COPD that I was trying to make her fit.* [P7]

Two students showed awareness of confirmation bias during the task and made a conscious effort to seek other information but consequently may have asked too many irrelevant questions and led to students taking a thorough approach to gathering data:

> *I think I normally like to sort of focus on a system, so do almost a respiratory thing and then move on to cardiac...although I shouldn't get too into confirming about, I'll just ask about any other symptoms.* [P4]
>
> *Often I ask confirmatory questions...and just exclude things that I just know weren't on my differential and so my differential didn't really change.* [P11]

Reflection

A few students reported that they found it useful that eCREST gave them time to pause, think, and reflect on their diagnoses because in clinical practice this is not always possible. The opportunity to reflect may explain why students were able to demonstrate the use of focused data gathering strategies and avoid some cognitive biases in eCREST:

> *It's nice to just click the questions, and then spend five minutes thinking about it. I think when you're actually seeing patients there's emphasis on it being slick.* [P14]

Alternative Diagnoses

There was evidence that eCREST helped students to avoid anchoring and confirmation bias, as most students reported that the prompts in eCREST to reassess diagnoses helped them to consider alternative diagnoses and reflect on the information they were gathering. This may have helped students to take a more focused approach to gathering data:

> *The fact that it makes you reconsider...your diagnosis after asking questions, asking a set number of questions is good practice for reality, when you should be doing that but you probably don't.* [P10]
>
> *It makes you like re-evaluate your ranking of diagnosis because then you actually have think about the questions and why you're asking them in the first place.* [P10]

However, some students demonstrated a random approach to reassessing their diagnoses and reported that the prompts to revise their diagnoses did not always help them to consider alternative diagnoses and stay open minded.

> *I'm just going to put in, arbitrarily, asthma...ischemic heart disease.* [P8]

A few students felt that the process of continually reviewing their diagnoses was too structured and some forgot information they gathered earlier in the case, which may have explained why their data gathering strategy appeared random:

> *It did perhaps make me think in that kind of modular way...each time I only considered the six questions that had been before. And forget about what had happened before that. So, like less of a kind of continuous set of questions and more like, oh, in the last six questions she said she like didn't have a fever.* [P16]

Theme 4: Confidence and Uncertainty

The way in which eCREST influenced students' confidence and level of uncertainty in their clinical decision-making skills appeared to be another potential mechanism by which eCREST influenced how students gathered information.

Creating Uncertainty

Some students reported that the structure of eCREST created some uncertainty, as the lists of diagnoses and questions encouraged them to reconsider diagnoses and think of the reasons behind choosing a diagnosis. This again suggests that the prompts to review diagnoses and reflect might have helped students to avoid anchoring and have a focused approach to gathering information from the patient:

> *I liked that there were all the differentials, you thought you knew but you still, it put that sort of seed of doubt in your mind. So in a sense...even though maybe I wasn't as flawless in terms of safety netting and making sure I didn't miss things, I think it forced you think okay well could it? And...constantly, by making you reconsider your diagnosis it did really make you think harder about each one.* [P4]

Managing Uncertainty

Some students talked about needing more confidence in their decision making to help them manage their uncertainty when gathering information from the patient:

> *Maybe I should just have more confidence in saying that, okay this is what I think and it's still consistent you know.* [P12]

Students also recognized a tension between making confident judgments about the most likely diagnosis and fearing serious consequences for the patient if they missed a more serious diagnosis. This may explain why some students adopted a thorough approach to gathering information and potentially used this approach to manage their uncertainty. There was also evidence from students' reflections that they were starting to manage this uncertainty by taking a more focused approach to gathering data and using safety netting:

> *You can give every investigation and then be sure, but actually realistically when you're trying to think well I can't get every single blood test in the world, this is the initial management...at the moment I feel I have to be quite brave, because you think well what if I do miss something that's terrible? But then I*

> *suppose it's easy to think okay, well what are the absolute terrible things? Make sure that I don't miss those, so for example an x-ray would cover a lot of bases in a sense.* [P4]
>
> *Learning to be comfortable...with a degree of uncertainty is important to GP and several other specialties. I will need to learn from more experience how much safety netting and investigation for other possible differentials is appropriate.* [P3]

Some students felt that eCREST and simulations were a way to practice taking responsibility for decisions and managing their uncertainty. It was suggested that being more engaged with the simulation helped students to adopt more focused or thorough data gathering patterns and avoid the unpacking principle, as it made them feel more responsible for the *patient* and motivated them to not miss any serious conditions:

> *I think every time I do...cases like this, and certainly when I see real patients, like I'm on GP at the moment...I'm going to try and think about it, to approach it as though I was the GP seeing the patient alone, and you're their only point of care...and therefore fully responsible for them. Which forces you to really think carefully about differentials and things not to miss.* [P4]

Some students felt reluctant to make decisions they would not be responsible for in real life and were perhaps less engaged with the simulation and the opportunity to practice managing uncertainty:

> *I would probably review him in the week, oh wait, actually, I don't know. I want to consult with my senior first but it's hard to say. I mean just to be...I mean if, in real life given, if I don't have any senior to talk to and if I don't know whether he has any back pain just to be safe I would probably just refer him, just to be sure, I guess.* [P13]

A few students felt that investigating rare and unlikely diagnoses was unrealistic and were perhaps less engaged with the simulation. This may have led to some students adopting more succinct data gathering approaches that just focused on common illnesses and led them to be more susceptible to cognitive biases:

> *I appreciate that...what they've said about all of their diagnoses but...in real life, what happens is: you make a working diagnosis, and everything else is left behind—you don't continue those, generally. There might be some things you safety net, but by and large, when it's clear cut—as that was—you would almost take that, go with it, do a few things, just to be sure. Where I am—and I know this is a GP situation, but in A&E, if you said, 'Oh, my fifth diagnosis of this patient is "Addison's Disease,"' I think you'd...be laughed out of the department, realistically.* [P1]

Identifying Data Gathering Patterns

The qualitative data showed that students aimed to apply 3 distinct data gathering strategies while using eCREST. It suggested that these strategies may manifest in the quantitative data as different data gathering behaviors and indicated how eCREST might influence them. As described in the *Methods* section, we sought to observe these data gathering patterns in the trial sample by using trial students' scores on the Essential and Relevant variables. Figure 2 summarizes the characteristics of the data gathering patterns. Those who displayed Focused and Thorough patterns tended to elicit the most essential information and take a more complete patient history. However, those who showed Thorough patterns also tended to gather more irrelevant information. Those who displayed a Succinct pattern or a Nonspecific pattern did not elicit enough essential information from the patient and took a less complete history from the patient. However, those who showed Succinct patterns also did not gather much irrelevant information.

Figure 3 shows a scatter plot of the trial students' scores on the Essential and Relevant variables by the trial group and whether their scores fell on or below the lowest quartile for each variable, which determined the data gathering pattern they were identified as displaying. The data show most students displayed Focused patterns, but a significant proportion displayed other patterns and the prevalence appeared to differ by the trial group.

Figure 2. Characteristics of data gathering patterns based on the “Essential” and “Relevant” variables.

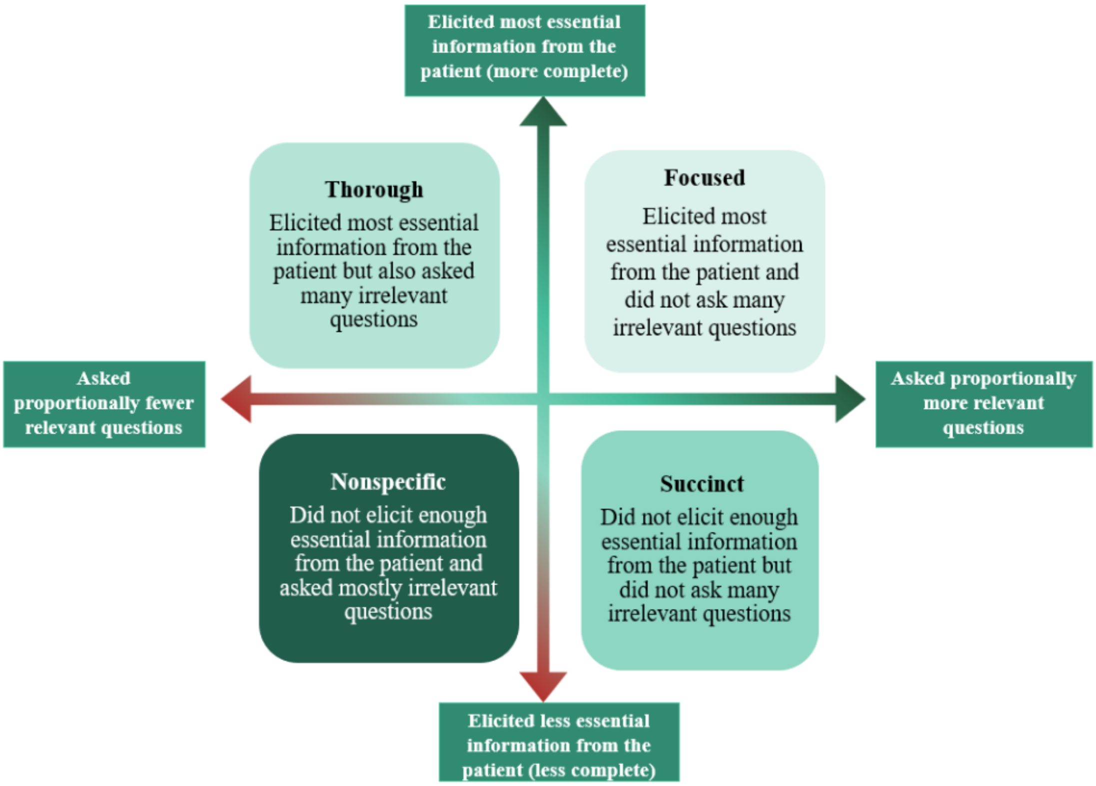
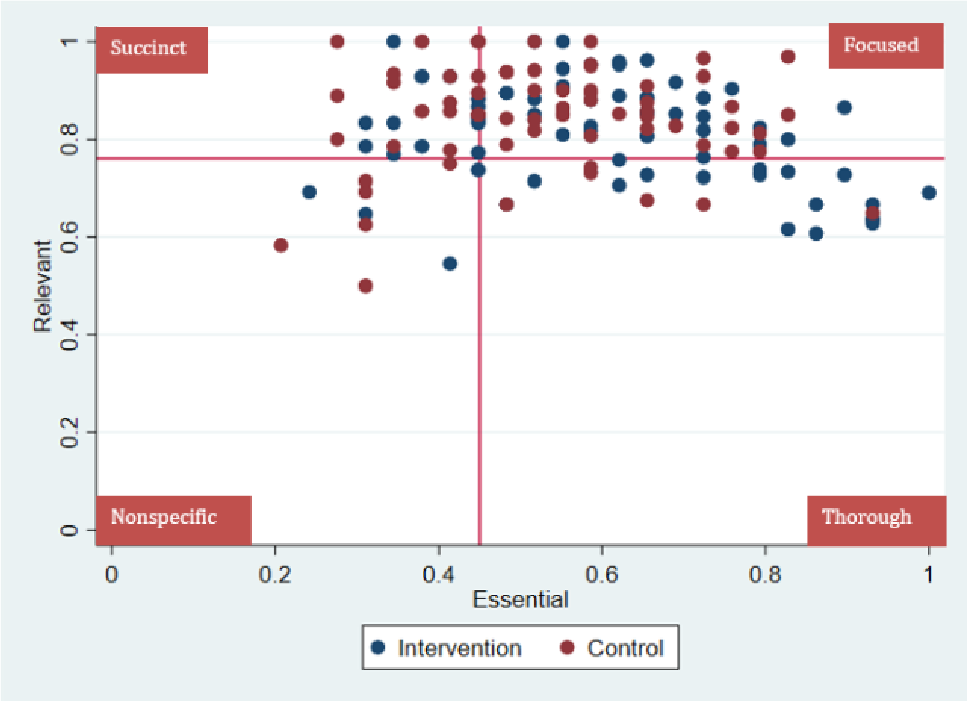


Figure 3. Scatter plot showing students’ scores on the “Essential” and “Relevant” clinical reasoning measure in the trial.



Impact of eCREST on Data Gathering Patterns

In the trial, there was a significant difference between the intervention and control groups in the type of data gathering pattern used (χ^2_3=9.9; *P*=.02; Table 2). Those in the intervention group were much more likely to show a Thorough pattern compared with the control group (21/78, 27% vs 6/70, 9%) but less likely to demonstrate a Succinct pattern (13/78, 17% vs 20/70, 29%). The likelihood of showing Focused or Nonspecific patterns were similar between the intervention and control groups (40/78, 51% vs 38/70, 54% and 4/78, 5% vs 6/70, 9%, respectively).

Table 2. Data gathering patterns observed in the electronic Clinical Reasoning Educational Simulation Tool in the trial data.

Data gathering pattern[a]	Intervention (n=78), n (%)	Control (n=70), n (%)	Total (N=148), n (%)
Thorough	21 (27)	6 (9)	27 (18.2)
Succinct	13 (17)	20 (29)	33 (22.3)
Focused	40 (51)	38 (54)	78 (52.7)
Nonspecific	4 (5)	6 (9)	10 (6.8)

[a] Patterns were significantly associated with the trial group; χ^2_3=9.9, *P*=.02 (as in the main text of the paper).

Discussion

Principal Findings

This study identified a range of data gathering patterns that students applied when generating data from virtual patients. The qualitative data indicated how eCREST can help students to take a more thorough approach, and potentially reduce the impact of cognitive biases, through continuous revision of diagnoses and allowing students to practice managing their uncertainty. Quantitative data from the trial indicated that eCREST influenced students to demonstrate more Thorough data gathering patterns.

This study showed how virtual patients such as eCREST can be used to address the cognitive biases of the unpacking principle, confirmation bias, and anchoring by continuously prompting students to reflect throughout a patient consultation. The qualitative data showed that these prompts encouraged many students to investigate patients more thoroughly and re-evaluate their diagnoses. This may have helped them to overcome the potentially negative consequences of these biases, such as missing serious diagnoses [32,47]. The trial data also indicated that students who had used eCREST before exhibited more Thorough data gathering patterns than controls, which may have helped them to address these biases [33]. We have no empirical data to suggest that any data gathering pattern is better than another and in which clinical circumstances they might be most appropriate. In clinical practice, a more Focused approach where most of the important information is gathered without gathering too much irrelevant information may be ideal, particularly given time constraints in real consultations. Policy and clinical guidelines in the United Kingdom and elsewhere are increasingly recommending more focused and thorough investigations of patients to avoid missing red flag symptoms, particularly in primary care for serious conditions such as cancer [8,25]. Thus, the data gathering pattern that eCREST is encouraging is in line with recommendations from health policy and may be particularly appropriate for the investigation of conditions such as cancer.

A unique contribution of this study is the use of a mixed methods approach. The qualitative data showed distinct data gathering strategies that students aimed to undertake and how eCREST influenced how they gathered information. We gained insight into students' rationales for their data gathering strategies. Those who reported wanting to be Thorough explained that they used this strategy to avoid missing key information about a serious diagnosis and because they felt uncertain about the case. The students who reported wanting to be more focused verbalized the importance of asking only relevant questions. Those who aimed to be more succinct reported wanting to limit the information they gathered. This was perhaps because in real consultations, students would only have limited time with a patient, but this may have led them to be susceptible to biases and miss symptoms indicative of a serious disease. We also found that some students were less engaged with eCREST and randomly clicked on questions and made decisions. Further research is needed to understand which students might not benefit from clinical reasoning teaching delivered via virtual patients and how it can be further adapted to students' needs.

Similar to previous literature on clinical reasoning, we identified a central theme for managing uncertainty [48,49]. The qualitative data showed how eCREST created uncertainty by prompting students to reconsider diagnoses. It also offered students an opportunity to practice managing their uncertainty by conducting thorough investigations and safety netting for the worst-case scenarios [48]. Therefore, eCREST may have helped to calibrate students' confidence. Previous studies have shown that there is little correlation between confidence and diagnostic accuracy in students or physicians and that overconfidence increases with more difficult cases [49-51]. Given that the cases in eCREST were relatively complex for students, it is perhaps a positive result that many students reported uncertainty and a lack of confidence while making diagnostic decisions in eCREST, suggesting that it might help reduce overconfidence in difficult cases.

This study also showed the potential impact of virtual patients, such as eCREST, on medical education. Given the increasingly limited exposure to real patients, virtual patients can provide students with some form of clinical experience [9]. They also provide an opportunity for educators to observe in real time students' data gathering strategies and patterns of behavior and inform the formative assessment of students' clinical reasoning skills. Previous research in education has used inventories to identify students' approaches to studying, such as the Approaches and Study Skills Inventory for Students [52-55]. Approaches and Study Skills Inventory for Students helps educators and students to identify general patterns on how students approach learning in certain circumstances and allows educators or computer programs to offer advice on other approaches that could be used. Similarly, educators could use data from eCREST in a reflective way and as a basis to provide feedback on data gathering patterns that might help students to improve their reasoning [21].

Limitations

Our study was limited to medical students, but we are undertaking further research with a range of health care professionals and students to understand how eCREST can be used more widely in clinical education. Students volunteered to take part in the trial and think-aloud study; therefore, this sample might have been different from those who did not volunteer, leading to possible selection bias. In addition, in the think-aloud study, students varied in their ability to verbalize their thoughts and the sample was not representative; therefore, it is likely that not all patterns and rationales for investigations were identified [56,57]. In common with all think-aloud and interview study designs, participants were observed and prompted to speak; therefore, students may have attuned their responses because of social desirability, and they were more reflective than they would have been if unprompted or unobserved [56,57].

Conclusions

This study found that students displayed a variety of data gathering patterns while using virtual patients. Data from the trial indicated that virtual patients such as eCREST might influence students to be more thorough in their data gathering. The think-aloud interviews suggested the mechanisms by which eCREST influenced students included helping them to continuously reflect on their diagnoses and manage uncertainty. These findings suggest that virtual patients, which increase the thoroughness of data gathering, could help future physicians to reduce missed diagnostic opportunities during future consultations. Virtual patients could also provide medical educators with a more accessible way of observing and identifying students' data gathering patterns, which may enable them to provide more tailored feedback on reasoning. Further research is needed to understand how data gathering patterns relate to existing clinical and pedagogical practices and vary across clinical contexts.

Acknowledgments

RP was supported by the Health Foundation for her PhD when she undertook this research and is currently supported by the National Institute for Health Research School for Public Health Research (grant PD-SPH-2015). JS was supported by the National Institute for Health Research Applied Research Collaboration North Thames. eCREST was commissioned and funded by the National Institute for Health Research Policy Research Programme, conducted through the Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis. The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care or its arm's length bodies or other government departments. The funders had no involvement in the review or approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart of the procedure for the trial.
[PNG File , 766 KB - jmir_v23i6e24723_app1.png]

References

1. Higgs J, Jones M, Loftus S, Christensen N. Clinical Reasoning in the Health Professions. Amsterdam: Elsevier; 2008:1-520.
2. Norman G. Research in clinical reasoning: past history and current trends. Med Educ 2005 Apr;39(4):418-427. [doi: 10.1111/j.1365-2929.2005.02127.x] [Medline: 15813765]
3. Cheraghi-Sohi S, Holland F, Reeves D, Campbell S, Esmail A, Morris R, et al. The incidence of diagnostic errors in UK primary care and implications for health care, research, and medical education: a retrospective record analysis of missed diagnostic opportunities. Br J Gen Pract 2018 Jun 06;68(suppl 1):bjgp18X696857. [doi: 10.3399/bjgp18x696857]
4. Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013 Oct;22 Suppl 2:21-27 [FREE Full text] [doi: 10.1136/bmjqs-2012-001615] [Medline: 23771902]
5. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med 2005 Jul 11;165(13):1493-1499. [doi: 10.1001/archinte.165.13.1493] [Medline: 16009864]
6. Hogan H, Healey F, Neale G, Thomson R, Vincent C, Black N. Preventable deaths due to problems in care in English acute hospitals: a retrospective case record review study. BMJ Qual Saf 2012 Sep;21(9):737-745 [FREE Full text] [doi: 10.1136/bmjqs-2011-001159] [Medline: 22927487]
7. Audétat MC, Dory V, Nendaz M, Vanpee D, Pestiaux D, Perron NJ, et al. What is so difficult about managing clinical reasoning difficulties? Med Educ 2012 Feb;46(2):216-227. [doi: 10.1111/j.1365-2923.2011.04151.x] [Medline: 22239335]
8. National Academies of Sciences, Engineering, and Medicine, Institute of Medicine, Board on Health Care Services, Committee on Diagnostic Error in Health Care. In: Balogh EP, Miller BT, Ball JR, editors. Improving Diagnosis in Health Care. Washington, DC: The National Academies Press; 2015:1-472.
9. Cleland JA, Abe K, Rethans J. The use of simulated patients in medical education: AMEE Guide No 42. Med Teach 2009 Jun;31(6):477-486. [doi: 10.1080/01421590903002821] [Medline: 19811162]

10. Greenhalgh T. Computer assisted learning in undergraduate medical education. Br Med J 2001 Jan 06;322(7277):40-44 [FREE Full text] [doi: 10.1136/bmj.322.7277.40] [Medline: 11141156]
11. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal. Med Educ 2015 Oct;49(10):961-973. [doi: 10.1111/medu.12775] [Medline: 26383068]
12. Guetterman TC, Sakakibara R, Baireddy S, Kron FW, Scerbo MW, Cleary JF, et al. Medical students' experiences and outcomes using a virtual human simulation to improve communication skills: mixed methods study. J Med Internet Res 2019 Nov 27;21(11):e15459 [FREE Full text] [doi: 10.2196/15459] [Medline: 31774400]
13. Hege I, Kononowicz AA, Adler M. A clinical reasoning tool for virtual patients: design-based research study. JMIR Med Educ 2017 Nov 02;3(2):e21 [FREE Full text] [doi: 10.2196/mededu.8100] [Medline: 29097355]
14. Edelbring S, Parodis I, Lundberg IE. Increasing reasoning awareness: video analysis of students' two-party virtual patient interactions. JMIR Med Educ 2018 Feb 27;4(1):e4 [FREE Full text] [doi: 10.2196/mededu.9137] [Medline: 29487043]
15. Padilha JM, Machado PP, Ribeiro A, Ramos J, Costa P. Clinical virtual simulation in nursing education: randomized controlled trial. J Med Internet Res 2019 Mar 18;21(3):e11529 [FREE Full text] [doi: 10.2196/11529] [Medline: 30882355]
16. Issenberg SB, McGaghie WC, Petrusa ER, Gordon DL, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005 Jan;27(1):10-28. [doi: 10.1080/01421590500046924] [Medline: 16147767]
17. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jul 02;21(7):e14676 [FREE Full text] [doi: 10.2196/14676] [Medline: 31267981]
18. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
19. Eva KW. What every teacher needs to know about clinical reasoning. Med Educ 2005 Jan;39(1):98-106. [doi: 10.1111/j.1365-2929.2004.01972.x] [Medline: 15612906]
20. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. Acad Emerg Med 2008 Nov;15(11):988-994 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00227.x] [Medline: 18778378]
21. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development. London: Prentice Hall; 1983:1-256.
22. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for Covid-19. Br Med J 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
23. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Health Ser Deliv Res 2018;6(21):A. [doi: 10.3310/hsdr06210] [Medline: 29939519]
24. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
25. The NHS long term plan. National Health Service. 2019. URL: https://www.longtermplan.nhs.uk/ [accessed 2021-05-17]
26. Silver District. URL: https://silverdistrict.uk/ [accessed 2021-05-21]
27. Kassianos A, Plackett R, Schartau P, Valerio C, Hopwood J, Kay N, et al. eCREST: a novel online patient simulation resource to aid better diagnosis through developing clinical reasoning. BMJ Simul Technol Enhanc Learn 2020 Jul;6(4):241-242 [FREE Full text] [doi: 10.1136/bmjstel-2019-000478] [Medline: 32832102]
28. Croskerry P. Achieving quality in clinical decision making: cognitive strategies and detection of bias. Acad Emerg Med 2002 Nov;9(11):1184-1204 [FREE Full text] [doi: 10.1111/j.1553-2712.2002.tb01574.x] [Medline: 12414468]
29. Kostopoulou O, Russo JE, Keenan G, Delaney BC, Douiri A. Information distortion in physicians' diagnostic judgments. Med Decis Making 2012;32(6):831-839. [doi: 10.1177/0272989X12447241] [Medline: 22622846]
30. Sheringham J, Sequeira R, Myles J, Hamilton W, McDonnell J, Offman J, et al. Variations in GPs' decisions to investigate suspected lung cancer: a factorial experiment using multimedia vignettes. BMJ Qual Saf 2017 Jun;26(6):449-459. [doi: 10.1136/bmjqs-2016-005679] [Medline: 27651515]
31. Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. Adv Health Sci Educ Theory Pract 2009 Sep;14 Suppl 1:27-35. [doi: 10.1007/s10459-009-9182-2] [Medline: 19669918]
32. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 2: impediments to and strategies for change. BMJ Qual Saf 2013 Oct;22 Suppl 2:65-72 [FREE Full text] [doi: 10.1136/bmjqs-2012-001713] [Medline: 23996094]
33. Plackett R, Kassianos AP, Kambouri M, Kay N, Mylan S, Hopwood J, et al. Online patient simulation training to improve clinical reasoning: a feasibility randomised controlled trial. BMC Med Educ 2020 Jul 31;20(1):245 [FREE Full text] [doi: 10.1186/s12909-020-02168-4] [Medline: 32736583]
34. Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. Med Educ 2003 Aug;37(8):695-703. [doi: 10.1046/j.1365-2923.2003.01577.x] [Medline: 12895249]
35. Johnson P, Duran A, Hassebrock F, Moller J, Prietula M, Feltovich P, et al. Expertise and error in diagnostic reasoning. Cogn Sci 1981;5(3):235-283. [doi: 10.1207/s15516709cog0503_3]
36. Tashakkori AM, Teddlie CB. SAGE Handbook of Mixed Methods in Social and Behavioral Research. Thousand Oaks, California, United States: Sage Publications Inc; 2010:1-912.

37. Caracelli VJ, Greene JC. Data analysis strategies for mixed-method evaluation designs. Educ Eval Policy Anal 2016 Nov 23;15(2):195-207. [doi: 10.3102/01623737015002195]
38. Johnson R, Onwuegbuzie A. Mixed methods research: a research paradigm whose time has come. Educ Res 2016 Jul 01;33(7):14-26 [FREE Full text] [doi: 10.3102/0013189X033007014]
39. Ericsson K, Simon H. Verbal reports as data. Psychol Rev 1980;87(3):215-251 [FREE Full text] [doi: 10.1037/0033-295X.87.3.215]
40. Ericsson K, Simon H. Protocol Analysis. Cambridge, Massachusetts, United States: The MIT Press; 1993:1-496.
41. Offredy M, Meerabeau E. The use of 'think aloud' technique, information processing theory and schema theory to explain decision-making processes of general practitioners and nurse practitioners using patient scenarios. Prim Health Care Res Dev 2005 Jan 01;6(1):46-59 [FREE Full text] [doi: 10.1191/1463423605pc228oa]
42. Johnson PE, Hassebrock F, Durán AS, Moller JH. Multimethod study of clinical judgment. Organ Behav Hum Perform 1982 Oct;30(2):201-230. [doi: 10.1016/0030-5073(82)90218-5] [Medline: 10257227]
43. NVivo 12. QSR International. URL: https://www.qsrinternational.com/ [accessed 2021-05-17]
44. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
45. Swain J. A hybrid approach to thematic analysis in qualitative research: using a practical example. SAGE Res Method Case Part 2 2018:A [FREE Full text] [doi: 10.4135/9781526435477]
46. Stata 15. Stata Statistical Software. URL: https://www.stata.com/stata15/ [accessed 2021-05-17]
47. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. Med Educ 2008 May;42(5):468-475. [doi: 10.1111/j.1365-2923.2008.03030.x] [Medline: 18412886]
48. Adams E, Goyder C, Heneghan C, Brand L, Ajjawi R. Clinical reasoning of junior doctors in emergency medicine: a grounded theory study. Emerg Med J 2017 Feb;34(2):70-75. [doi: 10.1136/emermed-2015-205650] [Medline: 27340131]
49. Balla J, Heneghan C, Thompson M, Balla M. Clinical decision making in a high-risk primary care environment: a qualitative study in the UK. BMJ Open 2012;2:e000414 [FREE Full text] [doi: 10.1136/bmjopen-2011-000414] [Medline: 22318661]
50. Filho GB, Moura AS, Brandão PR, Schmidt HG, Mamede S. Effects of deliberate reflection on diagnostic accuracy, confidence and diagnostic calibration in dermatology. Perspect Med Educ 2019 Aug;8(4):230-236 [FREE Full text] [doi: 10.1007/s40037-019-0522-5] [Medline: 31290117]
51. Meyer AN, Payne VL, Meeks DW, Rao R, Singh H. Physicians' diagnostic accuracy, confidence, and resource requests: a vignette study. JAMA Intern Med 2013 Nov 25;173(21):1952-1958. [doi: 10.1001/jamainternmed.2013.10081] [Medline: 23979070]
52. Tait H, Entwistle N. Identifying students at risk through ineffective study strategies. High Educ 1996 Jan;31(1):97-116 [FREE Full text] [doi: 10.1007/BF00129109]
53. Entwistle N, Tait H, McCune V. Patterns of response to an approaches to studying inventory across contrasting groups and contexts. Eur J Psychol Educ 2000 Mar;15(1):33-48 [FREE Full text] [doi: 10.1007/BF03173165]
54. Abedin N, Jaafar Z, Husain S, Abdullah R. The validity of ASSIST as a measurement of learning approach among MDAB students. Procedia Soc Behav Sci 2013 Oct;90:549-557 [FREE Full text] [doi: 10.1016/j.sbspro.2013.07.125]
55. Samarakoon L, Fernando T, Rodrigo C. Learning styles and approaches to learning among medical undergraduates and postgraduates. BMC Med Educ 2013 Mar 25;13:42 [FREE Full text] [doi: 10.1186/1472-6920-13-42] [Medline: 23521845]
56. Charters E. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. Brock Edu J 2003 Jul 01;12(2):A. [doi: 10.26522/brocked.v12i2.38]
57. Cotton D, Gresty K. Reflecting on the think-aloud method for evaluating e-learning. Br J Educ Technol 2006 Jan;37(1):45-54 [FREE Full text] [doi: 10.1111/j.1467-8535.2005.00521.x]

Abbreviations

eCREST: electronic Clinical Reasoning Educational Simulation Tool

---

Edited by G Eysenbach; submitted 02.10.20; peer-reviewed by S Edelbring, A Kononowicz; comments to author 15.11.20; revised version received 18.03.21; accepted 03.04.21; published 04.06.21.

Please cite as:
Plackett R, Kassianos AP, Timmis J, Sheringham J, Schartau P, Kambouri M
Using Virtual Patients to Explore the Clinical Reasoning Skills of Medical Students: Mixed Methods Study
J Med Internet Res 2021;23(6):e24723
URL: https://www.jmir.org/2021/6/e24723
doi:10.2196/24723
PMID:34085940

Original Paper

The Effects of Log-in Behaviors and Web Reviews on Patient Consultation in Online Health Communities: Longitudinal Study

Qin Chen[1], MSc; Jiahua Jin[1], PhD; Tingting Zhang[1], PhD; Xiangbin Yan[1], PhD

School of Economics and Management, University of Science and Technology Beijing, Beijing, China

Corresponding Author:
Jiahua Jin, PhD
School of Economics and Management
University of Science and Technology Beijing
No 30 Xueyuan Road, Haidian District
Beijing, 100083
China
Phone: 86 18518985411
Email: jinjiahua@ustb.edu.cn

Abstract

Background: With the rapid development of information technology and web-based communities, a growing number of patients choose to consult physicians in online health communities (OHCs) for information and treatment. Although extant research has primarily discussed factors that influence the consulting choices of OHC patients, there is still a lack of research on the effects of log-in behaviors and web reviews on patient consultation.

Objective: This study aims to explore the impact of physicians' log-in behavior and web reviews on patient consultation.

Methods: We conducted a longitudinal study to examine the effects of physicians' log-in behaviors and web reviews on patient consultation by analyzing short-panel data from 911 physicians over five periods in a Chinese OHC.

Results: The results showed that the physician's log-in behavior had a positive effect on patient consultation. The maximum number of days with no log-ins for a physician should be 20. The two web signals (log-in behavior and web reviews) had no complementary relationship. Moreover, the offline signal (ie, offline status) has different moderating effects on the two web signals, positively moderating the relationship between web reviews and patient consultation.

Conclusions: Our study contributes to the eHealth literature and advances the understanding of physicians' web-based behaviors. This study also provides practical implications, showing that physicians' log-in behavior alone can affect patient consultation rather than complementing web reviews.

(J Med Internet Res 2021;23(6):e25367) doi:10.2196/25367

KEYWORDS

online health communities; digital health; patient consultation; log-in behavior; web reviews; offline status

Introduction

Background

With the development of Health 2.0 technologies, the number of people using the internet to meet their health-related needs is increasing [1]. Online health communities (OHCs) have become prevalent in recent years, and research has focused on two of them: web-based physician-patient communities and web-based patient communities. A web-based physician-patient community is a platform that connects physicians with patients where patients can consult physicians on health issues and disease treatments anytime and anywhere. The object of this study is this kind of OHC, namely, the web-based physician-patient community.

Unlike other types of services, health care services have several characteristics. First, the disease of each patient is unique [2]. Second, life and death matter [3]. Third, serious information asymmetry exists between physicians and patients [4]. The choice of an appropriate physician has always been the focus of research in the health care field. The emergence of OHCs has effectively alleviated the problem of information asymmetry between physicians and patients. Unlike traditional health care services, OHCs give patients the opportunity to review the abundant amount of information about various physicians and then use this information to choose the physician whom they want to consult [5]. Although numerous studies have explored the factors that influence patients' choices of consulting [5-14], they are not the only ones; thus, more research is needed.

As health care service providers, physicians' web behaviors, such as knowledge-sharing behaviors [9], interactions with patients [5,6], and written and telephone consultations [13], provide important information when patients make consulting choices. However, for physicians, the premise of providing web-based health care services and conducting these behaviors involves logging into their accounts of the OHC. As physicians work full time in hospitals or clinics [14], they can only use off-duty hours to provide web-based services [15], which suggests that there are different log-in patterns in OHCs. Web-based behaviors reflect the degree of the physician's activeness and effort, as well as the quality of the service process [5,6,14,16]. In addition, as one kind of information generated by patients who have experienced health care services, web reviews can reflect the service outcome [14,16] and influence patients to make consulting choices [8-10]. The assessment of service quality should focus on both the outcome and delivery process they receive [17]. Log-in behavior and web reviews were evaluated in this study. Therefore, whether log-in behavior is related to patient consultation and its relationship with web reviews is worth studying.

As a web-based health platform may provide multiple signaling mechanisms simultaneously [18], there are multiple signals that influence patient consultation. This study focuses on both web-based and offline signals. According to the signaling theory [19], which indicates how the information receiver interprets signals along with information from the sender, log-in behaviors and web reviews can be considered as two web signals provided by a physician to assist patients in making consulting choices [6,11,20]. In addition, considering that people live in a mixed environment comprising web-based and offline worlds, the status of a physician's offline world (ie, offline status) as an offline signal may affect the relationship between web signals and patient consultation.

The objective of this study is to investigate how a physician's log-in behavior and web reviews affect patients' choices for consultation using data from a Chinese OHC. The main research questions are as follows:

1. How does a physician's log-in behavior affect patient consultation in OHCs?
2. How does a physician's log-in behavior and web reviews complement each other in affecting patient consultation?
3. How does the offline status of physicians moderate the effects of web signals (ie, log-in behavior and web reviews) on patient consultation?

To answer these three research questions, we collected data from 911 physicians over five periods and proposed a research model based on the signaling theory.

Related Research

With the development of information technologies, many physicians and patients are using OHCs. An OHC is a web-based community that presents a medical ecosystem, including patients and physicians, and is a core communication platform wherein patients can obtain physicians' web-based services, knowledge about diseases, and emotional support [20,21]. As health care service characteristics [2-4], the patient's choice of an appropriate physician for health care consultation has been the focus of research in the health care field. Table 1 summarizes the studies on patient consultation with OHCs. Although there are many factors that influence patients' choices of consulting [5-14], not all of them have been studied. More research is needed to better understand how information affects patient consultation in OHCs.

Table 1. Studies of patient consultation in online health communities.

Study	Theory	Influencing factors
Cao et al [5]	ELM[a]	The number of current patients who repeatedly interact with the physician, voting heating, service star, disease knowledge, and disease risk
Deng et al [6]	N/A[b]	Physician effort and web reputation
Li et al [7]	ELM	Technical quality, interpersonal quality, votes, high-privacy disease, and private doctor service
Li et al [9]	N/A	Web-based rating and activeness
Li et al [8]	N/A	Technical skills, interpersonal skills, and gender
Liu et al [11]	Signaling theory	The physician's web reputation and offline reputation; the hospital's web reputation and offline reputation
Liu et al [10]	N/A	Web-based service reviews, offline service reviews, and disease risk
Lu and Wu [12]	Service quality theory	Technical quality, functional quality, and disease risk
Wu and Lu [13]	N/A	Written consultation, telephone consultation, and doctor reputation
Yang et al [14]	Signaling theory	System-generated information and patient-generated information

[a] ELM: elaboration likelihood model.

[b] N/A: not applicable.

An OHC is a web-based platform wherein physicians can provide more types of health care services than offline hospitals or clinics [22], such as network consultation, phone consultation, and appointment registration. Furthermore, physicians can update their personal information, publish articles, respond to consultations, and manage patients. In the literature, many scholars have investigated physicians' web-based behaviors, such as knowledge-sharing behaviors [9,23,24], interactions with patients [5,6], written and telephone consultations [13], and contribution behaviors [25-27].

Although several web-based behaviors of physicians have been studied, not all of them have been evaluated. Log-in behavior is a web-based behavior of physicians that involves launching profile home pages in health websites and logging into their accounts. Log-in behavior is the first step for physicians to participate in OHCs and conduct other web-based behaviors. Many physicians offer services in both OHCs and hospitals or clinics. Owing to the heavy workload at hospitals or clinics (offline services), they can only use their spare time to provide patients with web-based services [15]. Thus, physicians have unique log-in behaviors. Previous research has shown that information about physicians' web-based behaviors is an important factor influencing patients' consultation choices [5,6,13]. However, less attention has been paid to log-in behaviors in OHCs. Therefore, the aim of this study is to explore physicians' log-in behaviors and their roles in OHC consulting choices.

Research Model and Hypotheses

Signaling Theory

The signaling theory is used to describe the behaviors of two parties (individuals or organizations) when accessing different information and has been applied in studies of investment decisions, entrepreneur-investor relationships [4], and web-based social trading [28]. The primary parties in the signaling theory include signalers and receivers as well as the signal itself. A signaler sends signals to the receiver to reflect quality [19]. The receiver evaluates the quality of the signaler and acts. As the two parties hold different amounts and levels of information, significant information asymmetry exists between the signalers and receivers [29]. Hence, the signal conveyed by signalers affects the degree of information asymmetry and can affect the receivers' behaviors.

Parties in OHCs include physicians and patients. Patients are at a disadvantage, as they must rely on physicians to provide health care services [30]. Physicians, as signalers, can provide information (eg, titles, workplaces, web-based behaviors, or reviews) to receivers (patients) [11], which can help patients choose physicians to serve their needs. Referring to past studies [11,14], this study used the signaling theory to explain the effects of log-in behavior and web reviews on patient consultation choices.

Log-in Behavior and Web Reviews

For patients, web-based behavior is often an important factor in choosing a physician. On the one hand, web-based behavior indicates the level of active participation that stems from internet motivation within the web-based community. Activeness has a certain influence on the number of patient consultations [9]. On the other hand, web-based behavior is a positive indicator of a physician's effort and popularity. Patients can gain insight into a physician's past efforts through web-based behaviors, which may influence their attitudes toward the physician, thereby influencing the likelihood of selecting that physician [6]. Most importantly, physicians' web-based behaviors are important cues for evaluating service process quality [14,16]. Combined with health care service characteristics [2-4], patients prefer to choose physicians who can provide a high-quality service process.

As the web-based behavior of a physician, log-in behavior reflects the degree of active participation in the OHC and the physician's efforts. Given that a physician may log in many times each day to check for new messages, log-in is the basis for any active actions for physicians in OHCs. Li et al [31] believed that log-in behavior belongs to the central working sphere, and log-in patterns could indicate a physician's central efforts related to the work, as well as the physician's attitude toward service provision, participation degree, and responsibility. Bitner et al [32] revealed that customers' perceptions of service depend on service providers' efforts, and their behaviors will raise the purchase intentions or continuous purchase intentions of customers [33], thus having a positive effect on marketing sales or organizational performance [34]. Physicians as providers of web-based health care services also apply to this phenomenon [9]. On the one hand, physicians with a higher-frequency log-in are more likely to make more task-related efforts to attract a greater number of patients, and subsequent patients would consider this for references. On the other hand, a higher-frequency log-in appears to be more responsive and involved than others [35], with those physicians logging in more frequently being more likely to ensure timeliness in service delivery, leading to attracting more patients. Therefore, the following hypothesis is proposed:

> *Hypothesis 1: a physician's log-in behavior has a positive effect on patient consultation.*

Web reviews are a particular type of user-generated content or electronic word-of-mouth, are the most important information source in customers' decision-making processes [36], and are deemed more successful in influencing customer behaviors than traditional marketing, information provided by products or service providers, or promotion messages from third-party websites [37-39].

OHCs provide a feedback channel where patients can express their views on the physician's service and share treatment experiences on the web. This information can help patients understand a physician's service quality at a minimal cost. In OHCs, web reviews are generated by patients who have experienced health care services. The more web reviews about a physician presented in the OHC, the more patients have selected the physician for consultation [16]. The web reviews generated by patients with similar experiences are more objective and credible signals than traditional information from acquaintances [40], which can increase other patients' trust in the physician and reduce perceived risks [41]. Web reviews are signals that reflect a physician's service outcome [14,16].

Positive web reviews mean a higher outcome quality of the physician, which has been shown to influence patients to make consulting choices [5,8,9,12,14].

The coexistence of log-in behavior and web reviews may complement each other in driving patient consultation. As service is delivered via the interaction between the service provider and the receiver, the assessment of service quality includes not only the delivery process but also the outcome [17]. As per the preceding discussion, log-in behavior may send web signals reflecting the service delivery process from the physicians themselves. A physician with a positive log-in behavior is usually associated with a positive attitude toward the consultation service. Furthermore, web reviews send another web signal from patients who have visited the physician before, which represents the service outcome. A physician with positive web reviews is usually associated with positive outcome quality. On the basis of the characteristics of health care services [2-4], patients judging a physician rely on two types of web signals: service process quality (ie, log-in behavior) and service outcome quality (ie, web reviews). Physicians with both high outcome quality and process quality are scarce resources [18], and the demand of these physicians on the platform should be large, so a large number of patients choose these patients. As a result, log-in behavior and web reviews should complement each other. From the preceding discussion, we propose the following hypothesis:

> *Hypothesis 2: a physician's log-in behavior and web reviews have a complementary relationship that affects patient consultation.*

Moderating Effect of Offline Status

Offline status reflects a physician's abilities and performance in providing health care services in hospitals or clinics [11,13], referring to a physician's career titles, ranking, and position in the hospital [11,42]. Such information can help patients evaluate a physician's offline competence [9].

In traditional health care services, patients can only judge a physician's ability through limited information. In the case of other factors being considered to be the same, patients tend to choose physicians with a higher status or professional titles [18]. To a certain extent, physicians with a high-level offline status might have a heavy workload but not enough time to contribute via the internet [18]. For this reason, the log-in behaviors of physicians with high-level offline statuses have a weaker impact on patient consultation. Therefore, the following hypothesis is proposed:

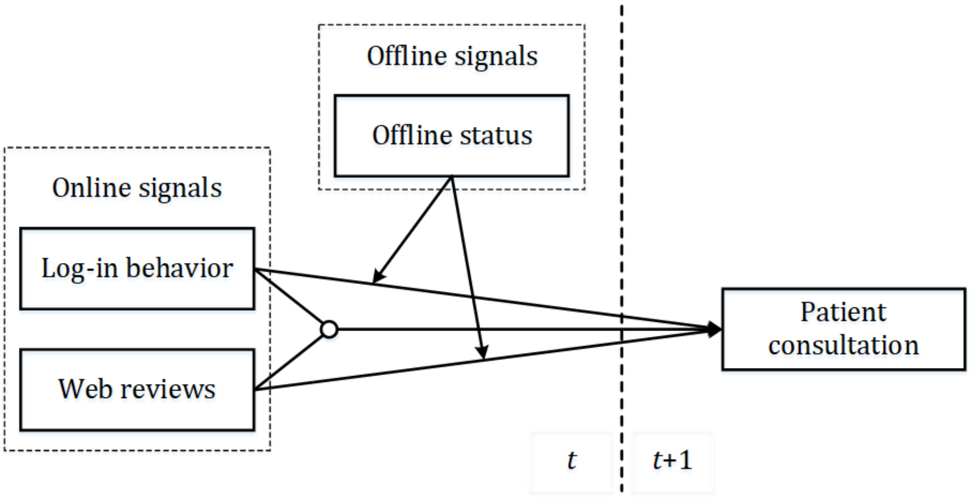
> *Hypothesis 3: the relationship between log-in behavior and patient consultation is negatively moderated by the physician's offline status.*

When patients choose physicians on the web, they rely mostly on offline and web signals. However, most patients regard offline signals as more reliable sources and treat web signals as additional information sources that supplement offline signals. Status may have a negative moderating effect on web signals [18]. When a physician has a high status, the offline signal will be sufficient for patients to make a decision. People are willing to accept the services of physicians with a high-level status as *credence services* instead of considering the service outcome reflected in web signals [25]. As physicians with high-level statuses may attract more patients, the effect of web reviews on patient consultation will be weakened. Hence, the following hypothesis is proposed:

> *Hypothesis 4: the relationship between web reviews and patient consultation is negatively moderated by a physician's offline status.*

The research model is shown in Figure 1.

Figure 1. Research model.



Methods

Research Context and Data Collection

The data used in this study were collected from Good Physician Online, which is one of the most popular and professional OHCs in China. It was founded in 2006 and, currently, more than 8000 hospitals and 500,000 physicians' information is presented on this website. Studying such a large and popular OHC can increase the generality of the results. Moreover, physicians registered on the Good Physician Online website have a profile home page, which contains information, such as physicians' background (name, medical title, academic title, hospital department, specialty, brief introduction, etc), patients' reviews, and information about web-based services. The information in a physician's profile home page is considerable and can help patients understand the physician and make a decision. Figure 2 shows an example of physician information shown on the OHC.

Figure 2. An example of physician information in online health communities.

综合推荐热度 4.4
在线服务满意度 100%
在线问诊量 6060

擅长 复杂疑难冠心病精准介入手术及综合治疗策略，复杂疑难心血管疾病的诊断和治疗，对心力衰竭，高血压，心律失常及心肌病等心血管疾病的诊断和药物治疗积累极其丰富的经验，在心血管急危重患者的诊治和 展开

简介 张铭，男，尔多斯人，主任医师，副教授，心血管内科硕士研究生导师，心血管内科医学博士，博士后，美国 Mayo Clinic postdoctoral research fellow，北京市卫生系统高层次人才心血管内科学科骨干， 展开

*基本信息来自医院官网或医院内公示信息或医生本人提供

欢迎词

医者仁心，你若相托，我必全力以赴！

非常感谢您的信任！郑重承诺：尽一生所学之医术，竭尽所能解决您的心血管疾病！尽管平时工作繁忙，但承诺24小时内网上100%回答您的各种问题。如患者确实需要住院行心脏病各类手术：如冠状动脉支架术、冠状动脉搭桥手术、心律失常导管消融手术和永久起搏器植入术及心脏瓣膜置换手术等，我们可以协助 展开

张铭的出诊时间

个人网站数据统计

总访问：	1,886,394次
昨日访问：	2,974次(2021-03-17)
总文章：	73篇
总患者：	6059位
昨日诊后报到患者：	3位
微信诊后报到患者：	1172位
总诊后报到患者：	1196位
患者投票：	280票
感谢信：	138封
心意礼物：	210个
上次在线：	今天
开通时间：	2012-02-13 14:25

张铭的患者投票 共280个

疗效满意度 99%　态度满意度 99%

疾病： 冠心病 (152) 心律失常 (11) 心衰 (8) 心脏病 (7) 高血压 (5) 房颤 (2)

患者：任***

看病目的：未填　治疗方式：介入手术　疗效满意度：很满意　态度满意度：很满意
本次看病费用总计：未填　目前病情状态：有好转　本次挂号途径：未填　选择该医生的理由：网上评价

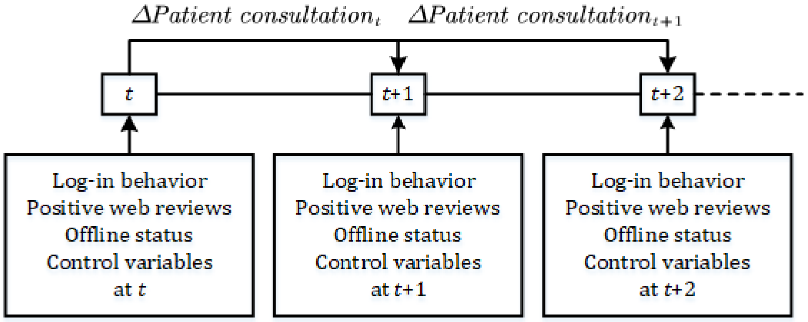
尊重病人 有微笑 问诊详细 理解病人焦虑 看诊中照顾情绪及痛苦 有解释也有安抚 好挂号 好住院 检查顺利

诊后服务星 ★★★☆☆
诊治后的患者数 3240例
随访中的患者数 964例

冠心病(502例) 心脏病(163例) 心律失常(140例) 高血压(124例) 心绞痛(76例) 心肌梗死(71例)

To reduce the influence of disease types, we only included physicians who treated patients with coronary heart disease as our sample. Using web crawler technology, we collected data from February 2019 to July 2019 (once every month during these six periods), which covered public information of hospitals and physicians presented on this website. We designed a longitudinal study to investigate whether a physician's log-in behavior and web reviews would change patient consultation choices. The data collection process is illustrated in Figure 3. After deletion of invalid data, short-panel data from 911 physicians over five periods were obtained for a total of 4555 physician data points. These physicians were currently active on the website, and the most recent log-in time was within 1 month.

Figure 3. Data collection and processing.



Variable Measurement

Table 2 presents the variable description. The dependent variable in this study was patient consultation. We used the number of patients before time *t* as a proxy for patient consultation, in accordance with previous research [5,14]. The number of patients included those who only consulted via the internet and those who consulted again after offline consultation. This study used the difference between the two periods as the dependent variable to reduce the causal relationship between the dependent and independent variables.

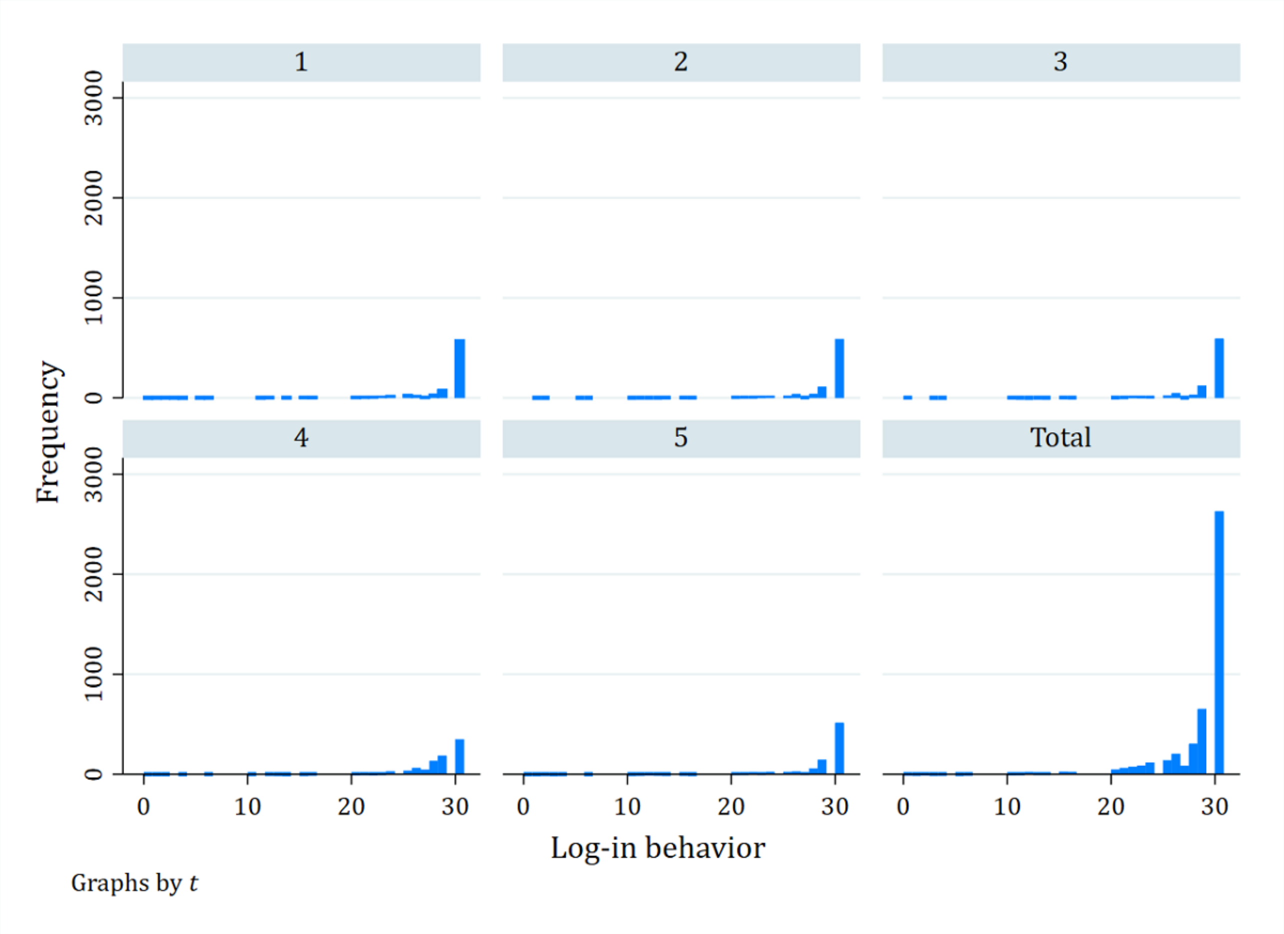
Table 2. Variable description.

Variables	Description	Proxy
Dependent variable		
Patient consultation	Patients evaluate the information about a physician and make the decision to consult online.	Patients
Independent variable		
Log-in behavior	One kind of online behavior that a physician launches his profile home page and log-ins his accounts.	Last web date
Positive web reviews	Positive reviews written by patients who have experienced a physician's health care service.	Thank-you letters
Moderating variable		
Offline status	The offline prestige of a physician in the career.	Medical title, academic title, or hospital ranking
Control variable		
Gender	The gender of a physician, 0 is male, 1 is female.	Gender
Usage years	The number of years that a physician using the OHC[a].	Usage years
Visits	The number of patients for visiting a physician’s profile home page.	Visits
Articles	The number of articles that a physician post on his or her profile home page.	Articles
Service stars	The number of service stars displayed on a physician’s profile home page.	Service stars
Written consultation	One type of online service. If a physician provided online written service, then 1; If not provided then 0.	Written consultation
Phone consultation	One type of online service. If a physician provided online phone service, then 1; If not provided then 0.	Phone consultation

[a] OHC: online health community.

The independent variables included physician’s log-in behavior and positive web reviews. In this study, log-in behavior was measured by the last date a physician was on the web at time *t*. If the physician logged in today, the log-in behavior is marked as 30; if the last on the web date is 30 days ago, the log-in behavior is marked as 0. The values are decremented individually. Figure 4 shows the frequency statistics of physicians’ log-in behaviors over five periods. It can be seen that physicians’ log-in patterns are different, and it is worth studying.

Figure 4. The frequency of physicians’ log-in behaviors in online health communities.



After the web-based consultation on the Good Physician Online website, patients can express their satisfaction or dissatisfaction with the physician’s service by sharing their treatment experience and writing a thank-you letter. The difference between a treatment experience and a thank-you letter is that the latter is a positive web review, and the patient who chooses to write a thank-you letter is definitely satisfied with the physician’s service. Referring to previous research [14,27], we used the number of thank-you letters a physician received from patients at time *t* to measure the physician’s positive web reviews.

The moderating variable in our research model is offline status, which mainly reflects the offline prestige in the physician’s career. According to previous research [27,42], we used the physician’s occupational title ranking and hospital standing at time *t* as a proxy for physicians’ offline status. A physician’s occupational title indicates the duties of a physician in a hospital, which is a manifestation of the physician’s professional expertise, health knowledge, and experience. A physician’s occupational title ranking includes medical title (chief physician, deputy chief physician, attending physician, and resident physician, coded from 4 to 1) and academic title (professor, deputy professor, and lecturer, coded from 3 to 1). Hospital standing can reflect an advantage in human capital, experience, health facilities, and technology, which is ranked as 3, 2, and 1. According to the methodology of previous research [43], this study integrated three variables to represent offline status. We standardized three variables by subtracting the means and dividing by the SEs, as shown in equation (1). Thus, the offline status of a physician was measured using equation (2).

Offline status = STD [STD (medical title) + STD (academic title) + STD (hospital ranking)] **(2)**

The control variables included the physician’s gender, usage years, number of visits, articles, service stars, written consultation, and phone consultation provided before time *t*. This information about the physician has been shown to be relevant to patients making consulting choices [5,7-9,13,14]. Gender is coded with “0” for male and “1” for female. Usage years are measured by the difference between the launching time of a physician’s personal website and time *t*. The number of visits, articles, and service stars is the information displayed on the physician’s profile home page before time *t*. Written consultation and phone consultation are two important types of services that physicians can provide on the Good Physician Online website. As the distributions of visits and articles are nonnormal, *ln* (*x*+1) transformations were also used for them.

Results

Descriptive Statistics and Correlation Results

Tables 3 and 4 show the descriptive statistics and correlations of variables, respectively. As shown in Table 4, log-in behavior is positively correlated with patient consultation and positive web reviews, and the β coefficients were .169 and .219, respectively. Log-in behavior and web reviews have positive correlations with offline status, with β coefficients of .047 and .284, respectively.

Table 3. Descriptive statistics of variables (n=4555).

Variables	Values		
	Mean (SD)	Min	Max
Gender	0.157 (0.364)	0	1
Usage years	5.127 (3.176)	0	11.340
Visits	718,657.100 (1,780,896.000)	27	1.80e+07
Articles	18.704 (72.534)	0	1314
Service stars	0.818 (1.217)	0	5
Written consultation	0.503 (0.500)	0	1
Phone consultation	0.643 (0.479)	0	1
Last web-based date	28.025 (4.105)	0	30
Thank-you letters	38.225 (93.953)	0	1606
Medical title	3.095 (0.850)	0	4
Academic title	1.186 (1.241)	0	3
Hospital ranking	2.940 (0.351)	0	3
Patients	18.013 (38.372)	0	678

Table 4. Correlations of variables (n=4555).

Variables	Gender	Usage years	ln(Visits+1)	ln(Articles+1)	Service stars	Written consultation	Phone consultation	Log-in behavior	Reviews	Status	Consultation
Gender	1	—[a]	—	—	—	—	—	—	—	—	—
Usage years	−0.056	1	—	—	—	—	—	—	—	—	—
ln(Visits+1)	−0.068	0.717	1	—	—	—	—	—	—	—	—
ln(Articles+1)	−0.150	0.321	0.502	1	—	—	—	—	—	—	—
Service stars	−0.041	0.082	0.223	0.188	1	—	—	—	—	—	—
Written consultation	0.021	−0.118	−0.203	−0.148	−0.068	1	—	—	—	—	—
Phone consultation	0.037	−0.152	−0.268	−0.195	−0.088	0.750	1	—	—	—	—
Log-in behavior	−0.035	0.066	0.131	0.109	0.259	−0.039	−0.037	1	—	—	—
Reviews	−0.088	0.448	0.647	0.292	0.498	−0.128	−0.174	0.219	1	—	—
Status	0.148	0.443	0.388	0.117	0.097	−0.029	−0.050	0.047	0.284	1	—
Consultation	−0.032	0.125	0.288	0.210	0.562	−0.020	−0.048	0.169	0.455	0.168	1

[a] Not applicable.

Estimation Model

As can be seen from Table 3, the dependent variables (patients) were nonnegative integers and their variance was greater than the mean; therefore, the negative binomial regression model was suitable for this study. The negative binominal probability function is as shown in equation (3), which has two parameters, θ and λ. Parameter θ captures overdispersion in the data, and parameter λ is the expected value of the distribution.

To test the hypotheses, the negative binomial regression model with fixed effects is explicitly expressed as shown in equation (4).

$$\triangle Patient\ consultation = Patient\ consultation_{i,t+1} - Patient\ consultation_{i,t}$$

$$= \alpha_0 + \alpha_1 Gender_i + \alpha_2 Usage\ years_{i,t} + \alpha_3 ln(Visits_{i,t} + 1) + \alpha_4 ln(Articles_{i,t} + 1) + \alpha_5 Service\ stars_{i,t} +$$

$$\alpha_6 \mathit{Written\ consultation}_{i,t} + \alpha_7 \mathit{Phone\ consulatation}_{i,t} + \alpha_8 \mathit{Log\text{-}in\ behavior}_{i,t} + \alpha_9 \mathit{Positive\ web\ reviews}_{i,t} + \alpha_{10} \mathit{Service\ stars}_{i,t} \times \mathit{Positive\ web\ reviews}_{i,t} + \alpha_{11} \mathit{Offline\ status}_{i,t} + \alpha_{12} \mathit{Log\text{-}in\ behavior}_{i,t} \times \mathit{Offline\ status}_{i,t} + \alpha_{13} \mathit{Positive\ web\ reviews}_{i,t} \times \mathit{Offline\ status}_{i,t} + \varepsilon_{i,t} \quad \mathbf{(4)}$$

Let i=1, 2, 3,..., n be the index of physicians. For equation (4), α_0 to α_{13} are the parameters to be estimated.

Regression Results

This study estimated the models using STATA software version 15.0 (StataCorp). The result of the Hausman test (χ^2_{14}=534.0; P<.001) indicated that the fixed effects model was suitable for this study. Table 5 shows the results of the fixed effects model hierarchically. Model 1 contains only constant and control variables, and model 2-model 5 add independent variables and interaction terms.

Table 5. Regression results (fixed effects model).

Variable	Model 1		Model 2		Model 3		Model 4		Model 5	
	α[a] (SE)	*P* value	α (SE)	*P* value	α (SE)	*P* value	α (SE)	*P* value	α (SE)	*P* value
Constant	−1.005 (0.324)	.002	−1.449 (0.344)	<.001	−1.059 (0.474)	.02	−1.074 (0.357)	.003	−.280 (0.347)	.42
Gender	−.269 (0.098)	.006	−.266 (0.098)	.007	−.241 (0.097)	.01	−.367 (0.101)	<.001	−.338 (0.100)	.001
Usage years	.034 (0.016)	.04	.033 (0.016)	.04	.035 (0.016)	.03	.015 (0.017)	.37	.016 (0.017)	.32
ln(Visits+1)	.199 (0.031)	<.001	.199 (0.031)	<.001	.135 (0.036)	<.001	.176 (0.031)	<.001	.110 (0.036)	.002
ln(Articles+1)	−.126 (0.027)	<.001	−.126 (0.027)	<.001	−.110 (0.027)	<.001	−.118 (0.027)	<.001	−.106 (0.027)	<.001
Service stars	.118 (0.015)	<.001	.112 (0.015)	<.001	.101 (0.015)	<.001	.111 (0.015)	<.001	.106 (0.015)	<.001
Written consultation	.104 (0.086)	.001	.106 (0.031)	.001	.106 (0.031)	.001	.107 (0.031)	.001	.104 (0.030)	.001
Phone consultation	−.302 (0.086)	<.001	−.311 (0.086)	<.001	−.303 (0.085)	<.001	−.323 (0.087)	<.001	−.319 (0.085)	<.001
Log-in behavior	—[b]	—	.016 (0.004)	<.001	.014 (0.011)	.20	.016 (0.004)	<.001	—	—
Positive reviews[c]	—	—	—	—	.105 (0.122)	.39	—	—	.128 (0.037)	.001
Log-in behavior×positive reviews	—	—	—	—	.001 (0.004)	.86	—	—	—	—
Offline status	—	—	—	—	—	—	.220 (0.142)	.12	−.023 (0.088)	.80
Log-in behavior×offline status	—	—	—	—	—	—	−.001 (0.005)	.80	—	—
Positive reviews×offline status	—	—	—	—	—	—	—	—	.070 (0.027)	.009
Log likelihood	−9349.410	—	−9341.811	—	−9336.253	—	−9331.336	—	−9329.970	—
Wald chi-square (*df*)	282.6 (7)	—	296.0 (8)	—	311.9 (10)	—	317.9 (10)	—	334.1 (10)	—
P value	<.001	—	<.001	—	<.001	—	<.001	—	<.001	—

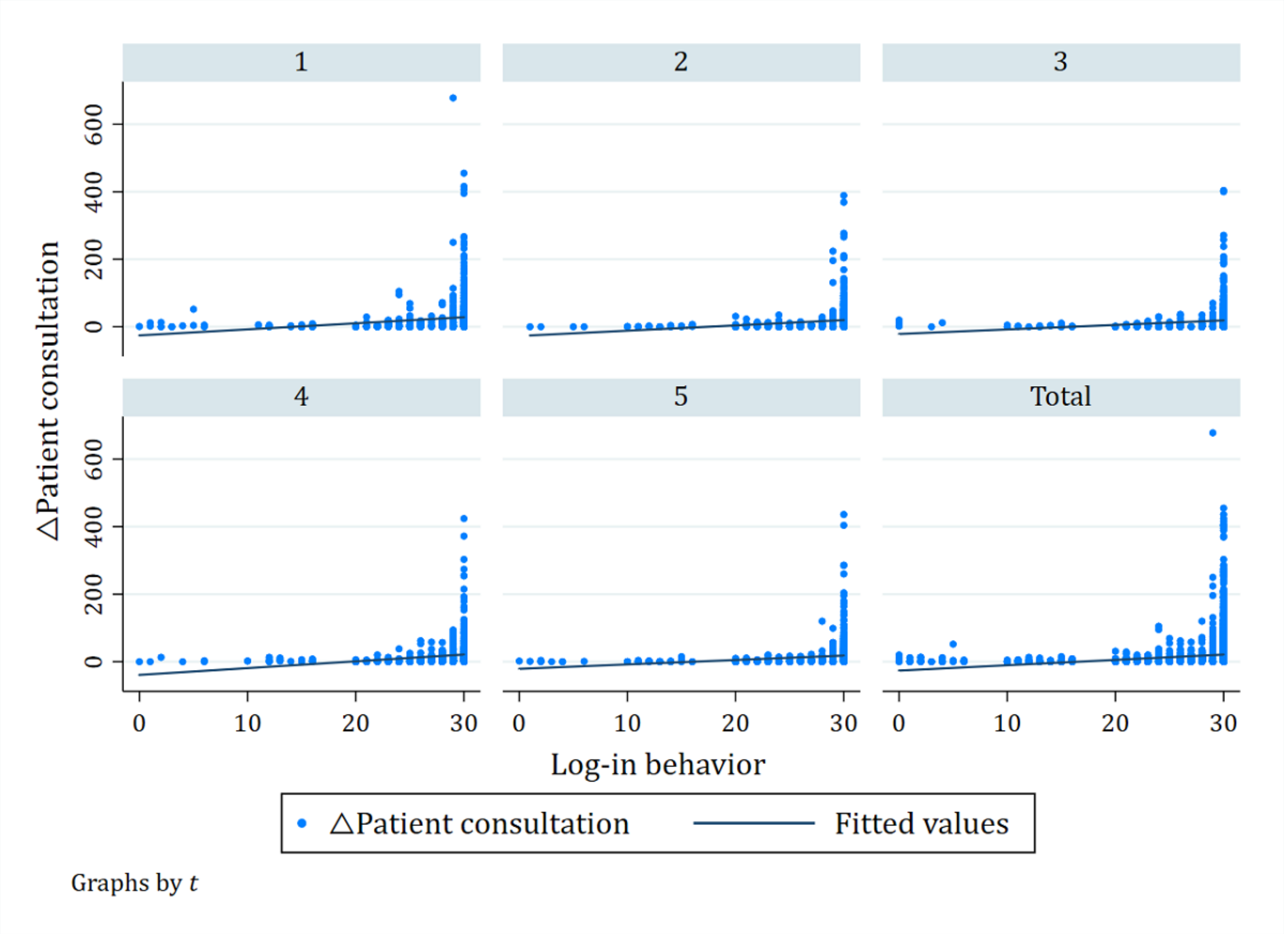
[a] Coefficient of the variable.

[b] Not applicable.

[c] Positive reviews: positive web reviews.

From model 2, the coefficient of log-in behavior (α=.016; P<.001) is positive and statistically significant, which supports hypothesis 1. The effects of log-in behavior on patient consultation are shown in Figure 5. As the number of log-in behaviors increases, the number of △patient consultation increases. When log-in behavior was less than 10 or 15, △patient consultation was 0 or <0. A comprehensive view of the regression lines of the five periods shows that the value of log-in behavior is ≤10, with the lowest number of patients making consulting choices.

Figure 5. The effect of log-in behavior and patient consultation on online health communities.



The results of model 3 show that the interaction between log-in behavior and web reviews (α=.001) is positive but not significant. This finding suggests that log-in behavior and web reviews do not have a complementary relationship that affects patient consultation. Therefore, hypothesis 2 is not supported.

The results of model 4 show that the interaction between log-in behavior and offline status (α=−.001) is negative but not significant. This means that the relationship between a physician's log-in behavior and patient consultation is not negatively moderated by offline status. Therefore, hypothesis 3 is contradicted.

The results of model 5 show that the interaction between web reviews and offline status (α=.070; *P*=.009) is positive and significant. This finding means that the effect of web reviews on patient consultation is stronger for physicians with a high status. Therefore, hypothesis 4 is contradicted.

Robustness Check

This study added the time effect to the estimation model, equation (5), and used the two-way fixed effects model to recheck the robustness of the results. Time is defined as a dummy variable, and $t1$ (February 2019) is used as the base period. The new estimation model is shown in equation (5). Table 6 shows the results of the robustness check, which are consistent with the results of the previous model (Table 5). In addition, the joint significance of the time dummy variable was tested, and it was confirmed that the time effect should be included in the estimation model. The robustness check results suggest that hypothesis 1 is supported.

$$\triangle Patient\ consultation = Patient\ consultation_{i,t+1} - Patient\ consultation_{i,t}$$

$$= \beta_0 + \beta_1 Gender_i + \beta_2 Usage\ years_{i,t} + \beta_3 ln(Visits_{i,t} + 1) + \beta_4 ln(Articles_{i,t} + 1) + \beta_5 Service\ stars_{i,t} + \beta_6 Written\ consultation_{i,t} + \beta_7 Phone\ consulatation_{i,t} + \beta_8 Log\text{-}in\ behavior_{i,t} + \beta_9 Positive\ web\ reviews_{i,t} + \beta_{10} Service\ stars_{i,t} \times Positive\ web\ reviews_{i,t} + \beta_{11} Offline\ status_{i,t} + \beta_{12} Log\text{-}in\ behavior_{i,t} \times Offline\ status_{i,t} + \beta_{13} Positive\ web\ reviews_{i,t} \times Offline\ status_{i,t} + \beta_{14} t_2 + \beta_{15} t_3 + \beta_{16} t_4 + \beta_{17} t_5 + \varepsilon_{i,t} \quad \textbf{(5)}$$

Let i=1, 2, 3,..., n be the index of physicians. For equation (5), β_0 to β_{17} were the parameters to be estimated.

Table 6. Robustness check (fixed effects model).

Variable	Model 1		Model 2		Model 3		Model 4		Model 5	
	β[a] (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value	β (SE)	*P* value
Constant	−.896 (0.338)	.008	−1.283 (0.356)	<.001	−.736 (0.466)	.11	−.966 (0.368)	.009	.141 (0.356)	.69
Gender	−.298 (0.101)	.003	−.294 (0.100)	.003	−.236 (0.098)	.02	−.392 (0.104)	<.001	−.329 (0.101)	.001
Usage years	.067 (0.018)	<.001	.065 (0.018)	<.001	.072 (0.017)	<.001	.046 (0.018)	.01	.055 (0.018)	.002
ln(Visits+1)	.219 (0.032)	<.001	.220 (0.032)	<.001	.084 (0.038)	.02	.200 (0.033)	<.001	.062 (0.038)	.10
ln(Articles+1)	−.090 (0.029)	.002	−.090 (0.029)	.002	−.056 (0.028)	.049	−.081 (0.029)	.006	−.056 (0.029)	.049
Service stars	.133 (0.014)	<.001	.128 (0.014)	<.001	.109 (0.014)	<.001	.128 (0.014)	<.001	.112 (0.014)	<.001
Written consultation	.002 (0.040)	.96	.003 (0.040)	.94	.001 (0.039)	.98	−.001 (0.040)	.97	−.001 (0.038)	.98
Phone consultation	−.185 (0.096)	.06	−.197 (0.096)	.04	−.178 (0.094)	.06	−.202 (0.097)	.04	−.196 (0.093)	.04
Log-in behavior	N/A[b]	N/A	.014 (0.004)	.001	.020 (0.010)	.06	.015 (0.004)	<.001	N/A	N/A
Positive reviews[c]	N/A	N/A	N/A	N/A	.346 (0.114)	.002	N/A	N/A	.278 (0.040)	<.001
Log-in behavior×positive reviews	N/A	N/A	N/A	N/A	−.003 (0.004)	.46	N/A	N/A	N/A	N/A
Offline status	N/A	N/A	N/A	N/A	N/A	N/A	.294 (0.135)	.03	−.141 (0.092)	.13
Log-in behavior×offline status	N/A	N/A	N/A	N/A	N/A	N/A	−.004 (0.004)	.36	N/A	N/A
Positive reviews×offline status	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	.109 (0.028)	<.001
T										
2	−.391 (0.024)	<.001	−.391 (0.024)	<.001	−.389 (0.024)	<.001	−.392 (0.024)	<.001	−.390 (0.023)	<.001
3	−.443 (0.033)	<.001	−.442 (0.033)	<.001	−.452 (0.032)	<.001	−.444 (0.033)	<.001	−.451 (0.031)	<.001
4	−.461 (0.025)	<.001	−.456 (0.025)	<.001	−.472 (0.024)	<.001	−.451 (0.025)	<.001	−.467 (0.024)	<.001
5	−.540 (0.025)	<.001	−.537 (0.025)	<.001	−.560 (0.025)	<.001	−.530 (0.025)	<.001	−0.552 (0.024)	<.001
Log likelihood	−9103.221	N/A	−9096.944	N/A	−9075.380	N/A	−9087.928	N/A	−9064.380	N/A
Wald chi-square (*df*)	1001.0 (11)	N/A	1015.4 (12)	N/A	1103.6 (14)	N/A	1040.8 (14)	N/A	1161.3 (14)	N/A
P value	<.001	N/A	<.001	N/A	<.001	N/A	<.001	N/A	<.001	N/A

[a] Coefficient of the variable.

[b] N/A: not applicable.

[c] Positive reviews: positive web reviews.

Discussion

Principal Findings

In contrast to previous studies on physicians' web-based behaviors, our research focused on log-in behavior and found that it had a positive effect on patient consultation. The results were consistent with those of Li et al [31], who believed that physicians with higher-frequency log-ins are more likely to attract patients, because they seem to be more responsible and have a timely service process. The results also indicated that physicians' web-based behaviors positively influence patients' consulting choices [6,9,13], including log-in behavior. Our research also found that when a physician did not log in to the OHC for more than 20 days, the number of patients who chose them was small, even 0.

Our research used web reviews generated by patients after receiving health care services as a web signal to represent service outcomes. Our research found that a physician's log-in behavior and web reviews did not have a complementary relationship in affecting patient consultation, which was different from the findings of previous research on service quality [12,14]. On the one hand, it may be that log-in behavior and web reviews have separate effects on patient consultation, and patients do not consider both. On the other hand, although log-in behavior is a web-based behavior, it may not be directly related to the delivery process of a physician's response to consultation.

Patients mostly rely on both offline and web signals to choose a physician. This study found that web reviews were positively moderated by offline status. This is inconsistent with the findings of previous research, which suggests that web signals should be negatively moderated by offline signals [18]. However, offline status cannot moderate log-in behavior. A possible explanation is that most patients view offline signals as a more reliable source than web signals. Compared with the degree of initiative and effort, offline prestige (ie, offline status) in a physician's career can better reflect the service outcome quality.

Theoretical Implications

This study offers theoretical contributions in the following ways. First, previous studies have explored the influencing factors related to patients' consultation choices, including some web-based behaviors of physicians, such as publishing articles, providing written consultation, and phone consultation. However, the literature on the role of physicians' log-in behavior is inadequate. Logging is the central working sphere and is the first step for a physician to provide health care services. Log-in behavior represents the central effort, activeness, and service process quality. Our research found that log-in behavior could influence patients' consulting choices. This finding extends the understanding of physicians' web-based behaviors and may also be used in other service fields.

Second, although some signaling literature in the context of eHealth has discussed web reviews, no research considers web reviews as service outcomes with the log-in behavior of physicians. However, this study found that log-in behavior and web reviews did not have a complementary relationship that affected patient consultation. Therefore, these findings contribute to research on patient consultation in OHCs.

Third, a clear distinction exists between web and offline signals. This study investigated the main effects of web signals (log-in behavior and web reviews) and their interactions with offline signals (offline status). The results revealed that the moderating effects of offline status on these two signals were different. From this perspective, this study extended the understanding of multiple signal interactions.

Implications for Practice

This study has several practical implications. First, for health care service providers, our evidence-based research demonstrates that log-in behavior is also an important factor in influencing patients' choice of consultation. Apart from other web-based behaviors, patients can judge a physician's activeness, efforts, and service process quality by relying on their log-in behavior. Physicians should value their web-based behaviors and log in to OHCs proactively, transmitting signals of active participation and timely responses to patients. Furthermore, operators of OHCs should pay attention to physicians' log-in issues. The more actively physicians participate in web-based platforms, the more successful the OHCs will be.

Second, the results show that log-in behavior and web reviews do not have a complementary relationship that affects patient consultation. Physicians should distinguish between log-in behavior and other web-based behaviors. Although web-based behaviors can reflect a physician's activeness and effort, there may be differences in service process quality.

Third, the results show that multiple signals from different signaling mechanisms affect patient consultation. Offline signals can have positive moderating effects on web signals. Hence, physicians should value the impacts of both web-based and offline service quality, and offline service quality is more credible than web-based service quality for patients.

Limitations and Future Research

This study has certain limitations. First, this study used physician data from only one OHC and one disease type. However, interpretation of the results may be limited. Therefore, it is necessary to collect data from physicians with various expertise on different platforms simultaneously to further verify the research model. Second, the study used the physician's last date displayed on the web on the day of crawling data to measure log-in behavior, which has certain limitations. In future research, we could measure log-in behavior through other methods, such as counting physicians' log-in times within a month. Third, the control variables selected in this study may have ignored some important variables, especially those related to patients. As websites tend to obscure customer names to protect privacy, it is difficult to obtain these data from the website.

Conclusions

Drawing on the signaling theory, this study explores the effects of physicians' log-in behavior and web reviews on patient consultation in OHCs. This study hypothesized that two signals (ie, log-in behavior and web reviews) and their interaction affect patients' consultation choices, and the relationships between web signals and patient consultation were moderated by offline signals (ie, offline status). Short-panel data over five periods were used to test these hypotheses. Our research found that a physician's log-in behavior positively affects patient consultation, and a physician's no–log-in days should be no more than 20 days. Log-in behavior and web reviews had no complementary relationship that affects patient consultation. Furthermore, offline status could only positively moderate web reviews instead of log-in behavior.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (71531013, 71729001, and 71601100), the Humanity and Social Sciences Foundation of Ministry of Education for Young Scholars (20YJC630214), and the Fundamental Research Funds for the Central Universities (FRF-TP-20-045A2, FRF-TP-20-022A1).

Conflicts of Interest

None declared.

References

1. Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective. J Med Internet Res 2017 Apr 24;19(4):e130 [FREE Full text] [doi: 10.2196/jmir.6834] [Medline: 28438725]
2. Marx GT. Soft surveillence: a growth of mandatory volunteerism in collecting personal information — "Hey Buddy Can You Spare a DNA?". FC 2006;2(1):33-46 [FREE Full text] [doi: 10.32855/fcapital.200601.004]
3. Fichman RG, Kohli R, Krishnan R. Editorial overview—the role of information systems in healthcare: current research and future trends. Inf Syst Res 2011;22(3):419-428. [doi: 10.1287/isre.1110.0382]
4. Akçura MT, Ozdemir ZD. A strategic analysis of multi-channel expert services. J Manag Inf Syst 2017 Apr 20;34(1):206-231. [doi: 10.1080/07421222.2017.1297637]
5. Cao X, Liu Y, Zhu Z, Hu J, Chen X. Online selection of a physician by patients: empirical study from elaboration likelihood perspective. Comput Human Behav 2017 Aug;73:403-412. [doi: 10.1016/j.chb.2017.03.060]
6. Deng Z, Hong Z, Zhang W, Evans R, Chen Y. The effect of online effort and reputation of physicians on patients' choice: 3-wave data analysis of China's good doctor website. J Med Internet Res 2019 Mar 08;21(3):e10170 [FREE Full text] [doi: 10.2196/10170] [Medline: 30848726]
7. Li C, Zhang E, Han J. Adoption of online follow-up service by patients: an empirical study based on the elaboration likelihood model. Comput Human Behav 2021 Jan;114:106581. [doi: 10.1016/j.chb.2020.106581]
8. Li S, Lee-Won RJ, McKnight J. Effects of online physician reviews and physician gender on perceptions of physician skills and Primary Care Physician (PCP) selection. Health Commun 2019 Oct 24;34(11):1250-1258. [doi: 10.1080/10410236.2018.1475192] [Medline: 29792519]
9. Li Y, Ma X, Song J, Yang Y, Ju X. Exploring the effects of online rating and the activeness of physicians on the number of patients in an online health community. Telemed J E Health 2019 Nov 23;25(11):1090-1098. [doi: 10.1089/tmj.2018.0192] [Medline: 30676279]
10. Liu F, Li Y, Ju X. Exploring patients' consultation behaviors in the online health community: the role of disease risk. Telemed J E Health 2019 Mar;25(3):213-220. [doi: 10.1089/tmj.2018.0033] [Medline: 29927721]
11. Liu X, Guo X, Wu H, Wu T. The impact of individual and organizational reputation on physicians' appointments online. Int J Electron Commer 2016 Jun 20;20(4):551-577. [doi: 10.1080/10864415.2016.1171977]
12. Lu N, Wu H. Exploring the impact of word-of-mouth about physicians' service quality on patient choice based on online health communities. BMC Med Inform Decis Mak 2016 Nov 26;16(1):151 [FREE Full text] [doi: 10.1186/s12911-016-0386-0] [Medline: 27888834]
13. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. Int J Med Inform 2017 Dec;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
14. Yang H, Guo X, Wu T, Ju X. Exploring the effects of patient-generated and system-generated information on patients' online search, evaluation and decision. Electron Commer Res Appl 2015 May;14(3):192-203. [doi: 10.1016/j.elerap.2015.04.001]
15. Lagu T, Hannon NS, Rothberg MB, Lindenauer PK. Patients' evaluations of health care providers in the era of social networking: an analysis of physician-rating websites. J Gen Intern Med 2010 Sep;25(9):942-946 [FREE Full text] [doi: 10.1007/s11606-010-1383-0] [Medline: 20464523]
16. Chen Q, Yan X, Zhang T. Converting visitors of physicians' personal websites to customers in online health communities: longitudinal study. J Med Internet Res 2020 Aug 26;22(8):e20623-e20709 [FREE Full text] [doi: 10.2196/20623] [Medline: 32845248]
17. Parasuraman A, Zeithaml VA, Berry LL. A conceptual model of service quality and its implications for future research. J Mark 1985;49(4):41. [doi: 10.2307/1251430]

18. Li J, Tang J, Jiang L, Yen DC, Liu X. Economic success of physicians in the online consultation market: a signaling theory perspective. Int J Electron Commer 2019 Mar 26;23(2):244-271. [doi: 10.1080/10864415.2018.1564552]
19. Connelly BL, Certo ST, Ireland RD, Reutzel CR. Signaling theory: a review and assessment. J Manage 2010 Dec 20;37(1):39-67. [doi: 10.1177/0149206310388419]
20. Wu H, Lu N. How your colleagues' reputation impact your patients' odds of posting experiences: evidence from an online health community. Electron Commer Res Appl 2016 Mar;16:7-17. [doi: 10.1016/j.elerap.2016.01.002]
21. Yan L, Tan Y. Feeling blue? Go online: an empirical study of social support among patients. Inf Syst Res 2014 Dec;25(4):690-709. [doi: 10.1287/isre.2014.0538]
22. Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. Int J Med Inform 2018 Feb;110:77-89. [doi: 10.1016/j.ijmedinf.2017.11.009] [Medline: 29331257]
23. Yan Z, Wang T, Chen Y, Zhang H. Knowledge sharing in online health communities: a social exchange theory perspective. Inf Manag 2016 Jul;53(5):643-653. [doi: 10.1016/j.im.2016.02.001]
24. Zhang X, Liu S, Deng Z, Chen X. Knowledge sharing motivations in online health communities: a comparative study of health professionals and normal users. Comput Human Behav 2017 Oct;75:797-810. [doi: 10.1016/j.chb.2017.06.028]
25. Liu G, Zhou L, Wu J. What affects patients' online decisions: an empirical study of online appointment service based on text mining. In: Proceedings of the International Conference on Smart Health. 2018 Presented at: International Conference on Smart Health (ICSH); July 1–3, 2018; Wuhan, China p. 204-210. [doi: 10.1007/978-3-030-03649-2_20]
26. Wang J, Chiu Y, Yu H, Hsu Y. Understanding a nonlinear causal relationship between rewards and physicians' contributions in online health care communities: longitudinal study. J Med Internet Res 2017 Dec 21;19(12):e427 [FREE Full text] [doi: 10.2196/jmir.9082] [Medline: 29269344]
27. Yang H, Du HS, He W, Qiao H. Understanding the motivators affecting doctors' contributions in online healthcare communities: professional status as a moderator. Behav Inf Technol 2019 Oct 23;40(2):146-160. [doi: 10.1080/0144929x.2019.1679887]
28. Kromidha E, Li MC. Determinants of leadership in online social trading: a signaling theory perspective. J Bus Res 2019 Apr;97:184-197. [doi: 10.1016/j.jbusres.2019.01.004]
29. Stiglitz J. Information and the change in the paradigm in economics. New Front Econ 2004;92(3):27-67. [doi: 10.1017/CBO9780511754357.004]
30. Evans RG. Supplier-induced demand: some empirical evidence and implications. In: The Economics of Health and Medical Care. London: Palgrave Macmillan; 1974:162-173.
31. Li J, Deng Z, Evans RD, Liu S, Wu H. How doctors take initiatives in online healthcare communities. Ind Manag Data Syst 2020 Jun 30;120(7):1401-1420. [doi: 10.1108/imds-12-2019-0684]
32. Bitner MJ, Booms BH, Mohr LA. Critical service encounters: the employee's viewpoint. J Mark 1994 Oct;58(4):95-106. [doi: 10.2307/1251919]
33. Caruana A. Service loyalty. Eur J Mark 2002 Aug;36(7/8):811-828. [doi: 10.1108/03090560210430818]
34. Borucki CC, Burke MJ. An examination of service-related antecedents to retail store performance. J Organiz Behav 1999 Nov;20(6):943-962. [doi: 10.1002/(sici)1099-1379(199911)20:6<943::aid-job976>3.0.co;2-9]
35. Petrovčič A, Petrič G, Manfreda KL. The effect of email invitation elements on response rate in a web survey within an online community. Comput Human Behav 2016 Mar;56:320-329. [doi: 10.1016/j.chb.2015.11.025]
36. Ye Q, Law R, Gu B, Chen W. The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Comput Human Behav 2011 Mar;27(2):634-639. [doi: 10.1016/j.chb.2010.04.014]
37. Gretzel U, Yoo KH. Use and impact of online travel reviews. Inf Commun Technol Tour 2008:35-46. [doi: 10.1007/978-3-211-77280-5_4]
38. Yang J, Mai E. Experiential goods with network externalities effects: an empirical study of online rating system. J Bus Res 2010 Sep;63(9-10):1050-1057. [doi: 10.1016/j.jbusres.2009.04.029]
39. Zhang Z, Ye Q, Law R, Li Y. The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews. Int J Hosp Manag 2010 Dec;29(4):694-700. [doi: 10.1016/j.ijhm.2010.02.002]
40. Gao G, Greenwood BN, McCullough J, Agarwal R. A digital soapbox? The information value of online physician ratings. In: Proceedings of the Academy of Management Annual Meeting. 2012 Presented at: Academy of Management Annual Meeting; August 3-7, 2012; Boston, MA, USA p. 16354. [doi: 10.5465/AMBPP.2012.16354abstract]
41. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
42. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. J Manag Inf Syst 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
43. Bresnahan TF, Brynjolfsson E, Hitt LM. Information technology, workplace organization, and the demand for skilled labor: firm-level evidence. Q J Econ 2002 Feb 01;117(1):339-376. [doi: 10.1162/003355302753399526]

Abbreviations

OHC: online health community

Edited by R Kukafka; submitted 29.10.20; peer-reviewed by X Song, H Mehdizadeh, MDG Pimentel; comments to author 05.03.21; revised version received 18.03.21; accepted 25.04.21; published 03.06.21.

Please cite as:
Chen Q, Jin J, Zhang T, Yan X
The Effects of Log-in Behaviors and Web Reviews on Patient Consultation in Online Health Communities: Longitudinal Study
J Med Internet Res 2021;23(6):e25367
URL: https://www.jmir.org/2021/6/e25367
doi:10.2196/25367
PMID:34081008

Original Paper

Why #WeAreNotWaiting—Motivations and Self-Reported Outcomes Among Users of Open-source Automated Insulin Delivery Systems: Multinational Survey

Katarina Braune[1,2], MD; Katarzyna Anna Gajewska[3,4], PhD; Axel Thieffry[5], PhD; Dana Michelle Lewis[6], BA; Timothée Froment[7], MA; Shane O'Donnell[7], PhD; Jane Speight[8,9], PhD; Christel Hendrieckx[8,9], PhD; Jasmine Schipp[8,10], BSc; Timothy Skinner[8,10], PhD; Henriette Langstrup[11], PhD; Adrian Tappe[12], BSc; Klemens Raile[1], MD; Bryan Cleal[13], PhD

[1]Charité - Universitätsmedizin Berlin, Department of Paediatric Endocrinology and Diabetes, Berlin, Germany
[2]Berlin Institute of Health, Berlin, Germany
[3]#dedoc° Diabetes Online Community, Berlin, Germany
[4]Population Health Sciences, Royal College of Surgeons in Ireland, Dublin, Ireland
[5]Novo Nordisk Center for Biosustainability, Technical University of Denmark, Copenhagen, Denmark
[6]#OpenAPS, Seattle, WA, United States
[7]School of Sociology, University College Dublin, Dublin, Ireland
[8]The Australian Centre for Behavioural Research in Diabetes, Melbourne, Australia
[9]School of Psychology, Faculty of Health, Deakin University, Geelong, Australia
[10]Department of Psychology, University of Copenhagen, Copenhagen, Denmark
[11]Department of Public Health, Section for Health Services Research, University of Copenhagen, Copenhagen, Denmark
[12]AndroidAPS, Hamilton, New Zealand
[13]Diabetes Management Research, Steno Diabetes Center Copenhagen, Copenhagen, Denmark

Corresponding Author:
Katarina Braune, MD
Charité - Universitätsmedizin Berlin
Department of Paediatric Endocrinology and Diabetes
Augustenburger Platz 1
Berlin, 13353
Germany
Phone: 49 30616454
Email: katarina.braune@charite.de

Abstract

Background: Automated insulin delivery (AID) systems have been shown to be safe and effective in reducing hyperglycemia and hypoglycemia but are not universally available, accessible, or affordable. Therefore, user-driven open-source AID systems are becoming increasingly popular.

Objective: This study aims to investigate the motivations for which people with diabetes (types 1, 2, and other) or their caregivers decide to build and use a personalized open-source AID.

Methods: A cross-sectional web-based survey was conducted to assess personal motivations and associated self-reported clinical outcomes.

Results: Of 897 participants from 35 countries, 80.5% (722) were adults with diabetes and 19.5% (175) were caregivers of children with diabetes. Primary motivations to commence open-source AID included improving glycemic outcomes (476/509 adults, 93.5%, and 95/100 caregivers, 95%), reducing acute (443/508 adults, 87.2%, and 96/100 caregivers, 96%) and long-term (421/505 adults, 83.3%, and 91/100 caregivers, 91%) complication risk, interacting less frequently with diabetes technology (413/509 adults, 81.1%; 86/100 caregivers, 86%), improving their or child's sleep quality (364/508 adults, 71.6%, and 80/100 caregivers, 80%), increasing their or child's life expectancy (381/507 adults, 75.1%, and 84/100 caregivers, 84%), lack of commercially available AID systems (359/507 adults, 70.8%, and 79/99 caregivers, 80%), and unachieved therapy goals with available therapy options (348/509 adults, 68.4%, and 69/100 caregivers, 69%). Improving their own sleep quality was an almost universal motivator for caregivers (94/100, 94%). Significant improvements, independent of age and gender, were observed in self-reported glycated hemoglobin (HbA_{1c}), 7.14% (SD 1.13%; 54.5 mmol/mol, SD 12.4) to 6.24% (SD 0.64%; 44.7 mmol/mol, SD 7.0; *P*<.001), and time in range (62.96%, SD 16.18%, to 80.34%, SD 9.41%; *P*<.001).

Conclusions: These results highlight the unmet needs of people with diabetes, provide new insights into the evolving phenomenon of open-source AID technology, and indicate improved clinical outcomes. This study may inform health care professionals and policy makers about the opportunities provided by open-source AID systems.

International Registered Report Identifier (IRRID): RR2-10.2196/15368

(J Med Internet Res 2021;23(6):e25409) doi:10.2196/25409

KEYWORDS

diabetes; artificial pancreas; automated insulin delivery; open-source; patient-led; user-led; peer support; online communities; diabetes technology; digital health; mobile health; medical device regulation; motivation; sleep quality; do-it-yourself

Introduction

Background

Despite significant advances in health care, pharmaceuticals, and technological developments, type 1 diabetes remains a challenging chronic condition to manage, impacting life expectancy and diminishing quality of life [1-3]. Only a small proportion of people with type 1 diabetes achieve glycated hemoglobin (HbA_{1c}) levels below 7.0% (58 mmol/mol), as recommended by therapeutic guidelines to reduce the risk of long-term diabetes-related complications [4-6]. The complexity of diabetes self-management bears a high cognitive load and can cause distress in everyday life, with approximately 40% of people with type 1 diabetes reporting distress and/or depressive symptoms, particularly prevalent among adolescents and young adults [7-10].

In addition to optimizing glucose levels and variability, diabetes technologies have the potential to ease complex decision making and thereby reduce the cognitive and emotional burden of diabetes self-management. The latest advances in diabetes therapy combine sensors for continuous glucose monitoring and insulin pumps with computerized control algorithms, thereby enabling automated adjustments to insulin delivery in response to the user’s changing glucose levels. Automated insulin delivery (AID) systems, also known as *artificial pancreas* or *(hybrid) closed-loop* systems, are in various iterations of development and automaticity. Although a variety of commercial AID systems are under development, and some have recently become available in a limited number of countries, they are not universally available, accessible, or affordable.

To fill in the gap, open-source AID systems, also called *Do-It-Yourself* Artificial Pancreas Systems (DIYAPS), have been created by people with diabetes, in the web-based community behind the hashtag #WeAreNotWaiting, with instructions and codes for these systems available freely and widely via open-source platforms. Although anyone can access this, each user has to take responsibility to build their individual system and use it at their own risk. Initial observational studies have described significant improvements in glycemic outcomes in smaller cohorts of open-source AID users of all age groups, including children and adolescents whose caregivers build and maintain these systems on their behalf [11-15]. Further studies reported improved sleep quality and uninterrupted sleep, in particular, reduced burden of diabetes management, increased confidence in achieving diabetes management goals, increased energy, and reduced mood swings among open-source AID users [15]. An in-silico study of the AndroidAPS algorithm showed similar glycemic improvements and concluded that this algorithm is both safe and effective [16].

Despite the potential benefits of open-source AID systems, little is known about the reasons why people with diabetes initially chose to use this technology. It is important to determine the lessons to be learned from the #WeAreNotWaiting movement, especially for stakeholders involved in research and commercial product development and regulation, such as academia, industry, health care professionals, governance, and regulatory bodies.

Objectives

As part of the OPEN (Outcomes of Patients’ Evidence with Novel, Do-it-Yourself Artificial Pancreas Technology) Project, the aim of this study is to investigate motivational factors for building, using, and maintaining an open-source AID system among adults with diabetes (type 1, 2, and others) and caregivers of children and adolescents with diabetes, as well as their self-reported clinical outcomes, through a population-based survey [17].

Methods

Study Design and Participants

From November 2018 to March 2019, we conducted a web-based, cross-sectional survey titled *DIWHY* (Multimedia Appendices 1 and 2). The survey design was created by the patient-led OPEN consortium [17], in collaboration with open-source AID users, and piloted by a small number of them before the final release. The Checklist for Reporting Results of Internet E-Surveys was used to guide survey development [18]. The survey was approved by the Charité–Universitätsmedizin Berlin Ethics Committee (EA2/140/18). Participants were eligible if they were adults (aged >18 years), living with diabetes (type 1, 2, or other), or being caregivers of a child or an adolescent with diabetes using an open-source AID system.

Procedures

Participants were invited through public announcements on the OPEN Project website, in the Facebook groups *Looped* (>6000 members) and *AndroidAPS users* (>1800 members, November

2018), other regional subgroups on Facebook, and by public posts on Twitter using the hashtags #WeAreNotWaiting and #DIYAPS. All posts were organic, meaning there was no paid promotion or targeted advertising of posts on any platform. All participants gave their consent electronically. Participation was anonymous and voluntary; no financial or other compensation was provided. Participants were able to choose between 2 language options (English and German). There was a version for adults with diabetes and one for caregivers. Data were collected and managed using secure Research Electronic Data Capture electronic data capture tools hosted at Charité [19].

Measures

Initial questions focused on demographics, the type of open-source AID systems used, estimated commencement date, and 3 HbA_{1c} values each preinitiation and postinitiation of open-source AID (self-reported for adults; for caregivers, their child's). In addition, participants were asked to provide their or their child's average time in range (TIR; sensor glucose 70 mg/dL/4.0 mmol/L-180 mg/dL/10.0 mmol/L) before and after the commencement of open-source AID.

Subsequently, participants' motivation to build an open-source AID was assessed with a single question: "What motivated you to build a Do-It-Yourself Artificial Pancreas system for yourself? Indicate your level of agreement with each statement." A total of 14 fixed-choice statements followed to conclude the stem "I built a DIYAPS..." (eg, "...to achieve better glycemic control," "...to improve my own sleep quality"). For each statement, a 5-point Likert-type scale was used (*fully applies* to *does not apply at all*). In addition, participants could indicate further motivational factors using free text.

Quantitative Analysis and Statistical Testing

To ensure the reporting of robust parameters regarding HbA_{1c} levels, entries with more than one missing HbA_{1c} value either before or after open-source AID implementation were not considered in the calculation of arithmetic means, SDs, and statistical tests related to HbA_{1c}. The reduction in the average HbA_{1c} levels before and after open-source AID implementation was assessed using the Wilcoxon signed-rank test (*P* value threshold of .05, paired: *TRUE*, and alternative hypothesis: *greater*). Entries not providing TIR values before and after open-source AID implementation were not considered for the computation of TIR-related descriptive statistics and testing for the increased TIR after open-source AID implementation (same statistical test as for HbA_{1c}, with alternative hypothesis set to *lower*). Quantitative analyses were conducted within the R programming framework (v4.0.2; R Core Team), and the ggplot2 package was used to generate figures.

Content Analysis

Content analysis was performed to analyze responses to open-ended questions [20]. A total of 3 researchers coded data and analyzed the responses thematically in 2 rounds. After the first round, which was open, inductive, and independent, 3 lists of codes were merged and combined into a final version. The second round of coding was deductive, and each of the coders assessed the content according to the final list of codes. The interrater reliability (percentage agreement for multiple raters) method was used to calculate the level of agreement between coders, and the final list of the most frequently discussed codes was generated [21]. Codes were then compared with assess the level of similarity, for example, an interrater reliability result of 100% indicated that all codes generated by individual coders matched.

Results

Characteristics of the Study Cohort

A total of 1125 individuals participated in the *DIWHY* survey. After excluding 25.6% (288/1125) incomplete responses, data from 897 individuals over 35 countries were analyzed. Detailed demographic characteristics are shown in Table 1. Participants were mostly from Europe (691/897, 77%), whereas 14% (125/897) were from North America, and 9% (78/897) were from other continents. Most adults (599/722, 82.9%) and caregivers (153/175, 87.4%) had a university degree or higher. Of the respondents, 26% (236/897) had a professional background in information technology and 19% (170/897) in biomedicine or health care. Furthermore, 82% (736/897) of the participants reported out-of-pocket expenses, with an average of US $530 and a maximum of US $1000 per year. In both groups, various types of open-source AID systems were used regularly, with Loop being the most popular system in North America and AndroidAPS being the most frequently used system in Europe. Otherwise, the geographical location and household income did not indicate any specific patterns.

Table 1. Participants' demographic and self-reported clinical characteristics.

Participant demographics	Children and adolescents (n=175)	Adults (n=722)	Total (N=897)
People with diabetes, gender, n (%)			
Female	83 (47.4)	311 (43)	394 (43.8)
Male	92 (52.6)	411 (56.8)	503 (55.9)
Other	0 (0)	2 (0.3)	2 (0.2)
People with diabetes, average age, years (SD)	9.7 (4.0)	41.8 (11.8)	35.6 (16.7)
Type of diabetes, n (%)			
Type 1	174 (99.4)	714 (98.9)	888 (98.9)
Type 2	0 (0)	4 (0.6)	4 (0.4)
Other	1 (0.6)	4 (0.6)	5 (0.6)
Average duration of diabetes, years (SD)	5.1 (3.9)	25.2 (13.3)	21.4 (14.4)
Average duration of open-source AID[a] use, mean (SD)	10.3 (10.0)	10.0 (19.1)	10.1 (17.6)
Type of open-source AID used regularly, n (%)			
OpenAPS	42 (28.4)	104 (16.6)	146 (18.8)
AndroidAPS	71 (48)	380 (60.6)	451 (58.2)
Loop	42 (28.4)	179 (28.5)	221 (28.5)
Other[b]	5 (3.4)	39 (5)	44 (5.7)
Region, country of residence, n (%)			
Europe	130 (74.3)	561 (77.6)	691 (76.9)
Austria	3 (1.7)	23 (3.2)	26 (2.9)
Bulgaria	9 (5.1)	7 (1)	16 (1.8)
Czech Republic	12 (6.9)	9 (1.2)	21 (2.3)
Finland	8 (4.6)	10 (1.4)	18 (2)
Germany	46 (26.3)	363 (50.2)	409 (45.5)
The Netherlands	0 (0)	10 (1.4)	10 (1.1)
Spain	3 (1.7)	11 (1.5)	14 (1.6)
Sweden	8 (4.6)	3 (0.4)	11 (1.2)
The United Kingdom	23 (13.1)	99 (13.7)	122 (13.6)
Other[c]	14 (8)	35 (4.8)	49 (5.5)
North America	21 (12)	104 (13.9)	125 (13.4)
Canada	5 (2.9)	18 (2.5)	23 (2.6)
The United States	16 (9.1)	86 (11.9)	102 (11.3)
Asia	12 (6.9)	14 (2.9)	26 (2.9)
South Korea	12 (6.9)	10 (1.4)	22 (2.4)
Others[d]	0 (0)	4 (0.4)	4 (0.4)
Western Pacific	12 (6.9)	39 (5.4)	51 (5.7)
Australia	12 (6.9)	29 (4)	41 (4.5)
New Zealand	0 (0)	10 (1.4)	10 (1.1)
Africa	0 (0)	1 (0.1)	1 (0.1)
South Africa	0 (0)	1 (0.1)	1 (0.1)
Education: highest completed, n (%)			

Participant demographics	Children and adolescents (n=175)	Adults (n=722)	Total (N=897)
No or some high school	19 (10.9)	54 (7.6)	73 (8.1)
High school	16 (9.2)	67 (9.4)	58 (6.5)
University	111 (64.1)	449 (62.9)	627 (71.1)
Degree or diploma	21 (12.1)	61 (8.5)	82 (9.2)
Doctorate	21 (12.1)	89 (12.4)	110 (12.4)
Occupational status[e], n (%)			
Full time	101 (58.4)	486 (67.6)	587 (65.8)
Part time	55 (31.8)	114 (15.9)	169 (18.9)
Unemployed	10 (5.8)	6 (0.8)	16 (1.8)
Retired	0 (0)	38 (5.3)	38 (4.3)
Student	2 (1.2)	58 (8.1)	60 (6.7)
Other	5 (2.9)	17 (2.4)	22 (2.4)
Professional background[e], n (%)			
Medicine	24 (18.5)	102 (19.5)	126 (19.2)
Tech	35 (26.9)	137 (26.2)	172 (26.3)
Other	71 (54.6)	284 (54.3)	355 (54.4)
Household annual net income[e], US $, n (%)			
<20,000	19 (12)	87 (14.1)	106 (13.6)
20,000 to 34,999	12 (7.6)	60 (9.7)	72 (9.2)
35,000 to 49,999	19 (12)	88 (14.2)	107 (13.7)
50,000 to 74,999	33 (20.9)	138 (22.3)	171 (22.1)
75,000 to 99,999	24 (15.2)	84 (13.6)	108 (13.9)
>100,000	40 (25.9)	124 (20)	165 (21.2)

[a] AID: automated insulin delivery.

[b] xDrip, Nightscout, offline uploader for Medtronic 600 series, HAPP, and custom or own developments.

[c] Belgium, Croatia, Denmark, France, Greece, Hungary, Ireland, Italy, Lithuania, Luxembourg, Norway, Poland, Portugal, Russia, Slovakia, and Switzerland.

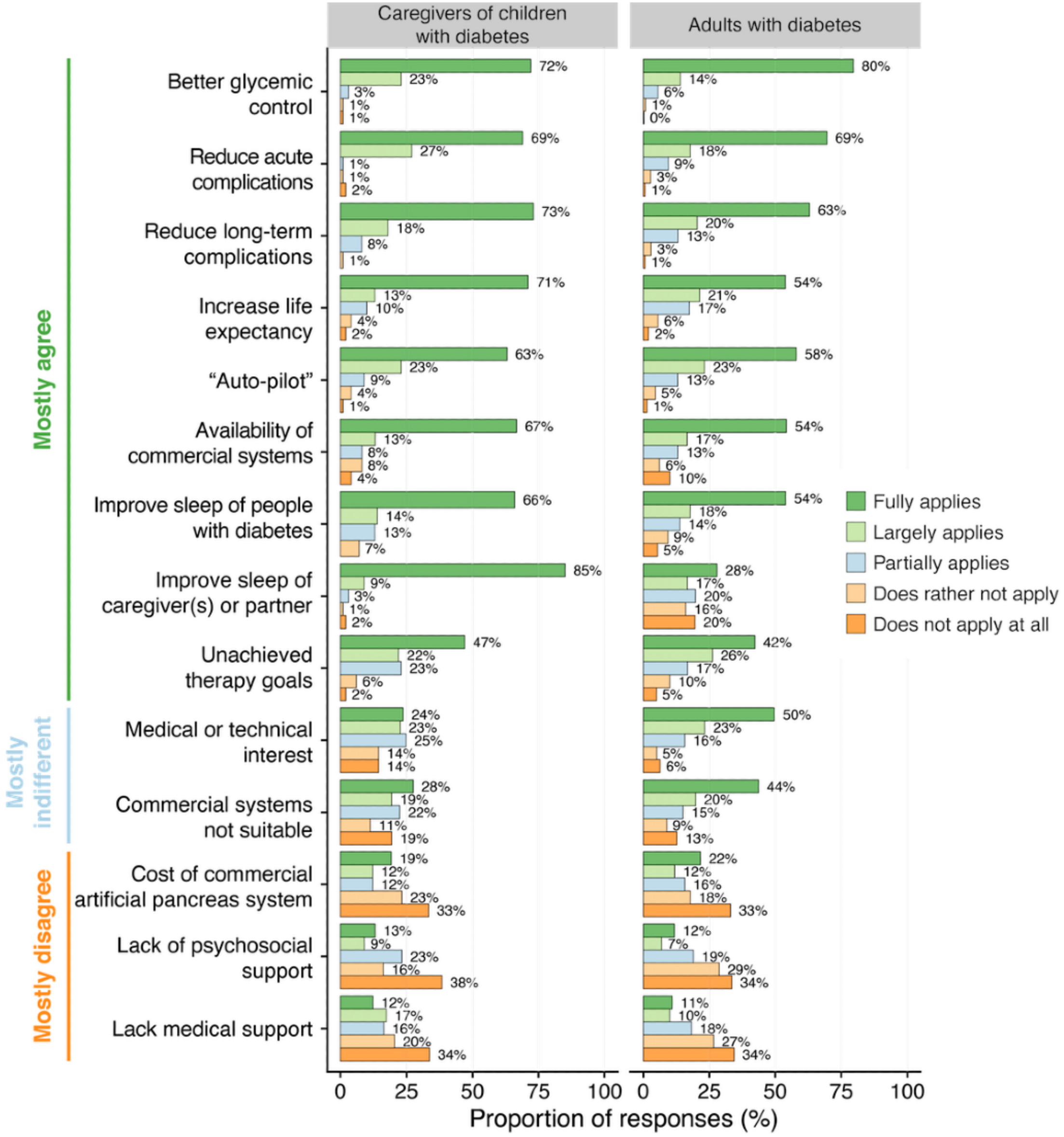
[d] Hong Kong, Kuwait, Palestine, and Singapore.

[e] For adults: own; for caregivers: caregivers.

Motivations to Commence Open-source AID Use

As shown in Figure 1, the most frequently endorsed motivations of adults as well as caregivers (as *fully applies* or *largely applies*) were to improve the overall glycemic control (476/509 adults, 93.5%; 95/100 caregivers, 95%), reduce the risk of acute (443/508 adults, 87.2%; 96/100 caregivers, 96%) and long-term complications (421/505 adults, 83.4%; 91/100 caregivers, 91%), put diabetes on *auto-pilot* mode and interact less frequently with diabetes technology (413/509 adults, 81.1%; 86/100 caregivers, 86%), increase their own or their child's life expectancy (381/507 adults, 75.1%; 84/100 caregivers, 84%), and improve their own or their child's sleep quality (364/508, adults 71.7%; 80/100 caregivers; 80%), because of the lack of commercially available closed-loop systems in their countries (359/507 adults, 70.8%; 79/99 caregivers, 80%) and unachieved therapy goals with the therapy options available to them (348/509 adults, 68.4%; 69/100 caregivers, 69%).

Figure 1. Motivations for building an open-source automated insulin delivery system. The x-axis shows the percentage of responses for each motivation question (y-axis). Bar colors represent the degree of relevance ranging from "does not apply at all" to "fully applies." The left and right columns show the responses of caregivers of children with diabetes and adults with diabetes, respectively. Responses are ranked from the most frequently endorsed motivations (top) to the less frequently endorsed (bottom).



Overall, the motivations of adults and caregivers of children and adolescents with diabetes were largely similar. As the most noticeable difference between the 2 groups, improvement in their own sleep quality (94/100, 94%) was a stronger motivation for caregivers compared with adults with respect to their partners or families (225/505, 44.6%). Curiosity (medical or technical interest) was endorsed more frequently by adults with diabetes (367/503, 73.0%) than by caregivers (45/97, 47%). Some believed that commercial systems did not suit their own or their child's individual needs, more frequently reported by adults (316/498, 63.5%) than by caregivers (46/98, 47%). Out-of-pocket costs related to the use of commercially available systems (166/496 adults, 33.5%; 31/99 caregivers, 31%) played a subordinate role. Lack of adequate medical support (105/501 adults, 21.0%; 29/98 caregivers, 30%) or psychosocial support (94/501 adults, 19.0%; 22/99 caregivers, 22%) were less frequently endorsed as motivating factors, although caregivers more frequently indicated a lack of medical support.

Further Motivations

In addition to the 14 predefined items, participants could indicate further motivation in an open-text field. In total, 127 participants (103 adults and 24 caregivers) provided a free-text response. Textbox 1 provides a list, as well as representative quotes, of the respondents. In the independent coders' selection of first-choice codes, there was an 83% interrater agreement between them (Multimedia Appendix 3).

Textbox 1. Illustrative quotes from adults with diabetes and the caregivers of children or adolescents with diabetes, highlighting additional motivation factors to build an open-source automated insulin delivery system.

Improving Diabetes Management

> *The psychological benefits of being able to significantly improve active control over diabetes and outcomes, rather than being more passively subjected to it.*
>
> *Another important reason for me is that I FINALLY have an overview of all data combined for later analysis but also direct decisions (values instead of opinions).*

Improving Quality of Life or Reducing the Burden of Diabetes Management

> *I chose DIY to decrease the demands of living with diabetes every day, around the clock. I also needed help consistently combatting the dawn phenomena, where I would wake up either too high, or too low from overcorrecting.*
>
> *His quality of life (staying with friends, knowing we can remotely monitor and assist, knowing that loop will help correct if he makes a mistake, attending sports training independently) is vastly improved. We can sleep! A happier, healthier family.*
>
> *Freedom to participate in normal 8 year old life eg play dates without having to pre-plan everything.*
>
> *Management of diabetes is helped by support but it is very much a self managed disease and requires 24/7 attention. Closed looping makes it just so much better, much of the time I can leave AAPS to take care of basals by itself. Quality of life is so much better. I can sleep without worrying about not waking up because of a bad hypo.*
>
> *We only wanted the best for our son. He should get exactly the same chances in life as his friends/children of the same age.*
>
> *To improve constant feeling of failure.*

Diabetes Distress or Burnout

> *Tired of diabetes after almost 30 years [...] The first real relief for me in my everyday life as a single mom.*
>
> *To reduce psychological distress, to be able to take responsibility for the course of diabetes, to enjoy life more since you are not torpedoed by Hypos and Hypers. Freedom despite technically higher dependency.*
>
> *There was no other way. The available treatments just did not control my diabetes sufficiently. The pressure and hopelessness of that scenario caused major mental health problems.*
>
> *I've lived my whole life [like] this & can't take it anymore. Too hard to do. Worst problem is "brain fog" & lack of energy due to blood sugar swings & hypoglycemic unawareness. I carry guilt for causing my family to lose sleep & carry the burden of diabetes [...]. [The] burden of diabetes is terrible.*

Autonomy

> *I feel so empowered by building my own system and taking control of my T1D. It's an awesome feeling!*
>
> *Daughter has learning difficulties, to make life easier for her and be less dependent on support, which in turn allows her to live a more independent life.*
>
> *To regain a sense of control on my diabetes management. I felt I was becoming dependent on my specialist for interpreting the adjustments needed for my insulin regime.*
>
> *Independent sleepovers with friends (without parents).*
>
> *To expand our daughter's independence and make her therapy decisions easier.*

Dissatisfaction With Available Technology, Choice and Health Care

> *Out of frustration with the existing designs seeming to have prioritized all stakeholders other than patients.*
>
> *Commercial closed loop systems do not allow users to specify a custom target BG but instead hard wire an unambitious target more concerned with legal liability that doesn't respect the autonomy, needs and wishes of the user.*
>
> *Dissatisfied with commercially available options and choice in the market space. No other option is appealing or provides the level of control and true artificial pancreas functions OR user interface.*
>
> *Doctors and hospitals have been telling me for years that things are simply fluctuating for me (hormones, stress, sensitivity to movement) and that you can't do anything about it. 'Resistant to all treatment options' and well-educated. Unfortunately with no success.*
>
> *We were desperate for something to use all the CGM data without sending our child crazy with in[sistent] requests for the pump to set low temps etc. We were infuriated by the business based decisions around closed loop in Australia - only the 640G was available and it was the worst decision for management and burden that we ever made. Now (as*

> *in within this fortnight) the 670G is available but still, no one can get training or sensors. We have been looping for nearly 3 years. If we hadn't then we would still be waiting today.*

Improving Sleep Quality

> *Sleep was the main reason followed by time in range. However, after all these years I still wake up but go back to sleep quickly.*
>
> *Frequently woke up from sensor alarms, make corrections and still wake up in the morning with a high or low glucose. Since closed looping, I get into bed knowing that Loop will keep me in range and I will wake up with a neat glucose. The only alarms I would ever get during the first period of closed looping were compression lows, and with the experience of loop keeping me in range I am now even confident enough to shut down all CGM alerts. Makes a huge difference for both me and boyfriend now that we start our days well rested. Every single day.*

Safety or Reducing Severe Hypoglycemia

> *Too many overnight hypos that require help.*
>
> *My child was overdosed on insulin twice by untrained teacher aides at school and if it was not for DIY looping technology- I would not have known about this at all until too late. Seeing the boluses appear on nightscout on real-time allowed me to question the dose and sugar treatment could commence preventatively than child actually going into severe hypo.*

"DIY mindset" or Early Adopter of Technology

> *I was going to build my own and found existing projects.*
>
> *Early adopter of all diabetes technologies. Turns disadvantage into a challenge.*
>
> *I love tinkering and making things. I'd always rather DIY, in many aspects of life.*
>
> *I initially built a closed DIY APS for a hackathon project out of pure tech curiosity. I planned to use the system for only 12 hours and then give a presentation to other employees at our company involved with the hackathon. After 12 hours, I realized I was never going to stop using it. Once on the system, almost every single one of the survey questions above are a "Fully Applies" as to why I decided to stay on the DIY APS.*
>
> *I'm a doctor and I'd like to test the closed loop first for myself and then use it in the future in my patients' treatment.*

Community Spirit

> *Being part of the community of selfless, generous, caring, and talented people willing to volunteer their time, knowledge, skills and experience to the benefit of the community.*
>
> *Something that also influenced me to move to a DIY system was the support from the community, and the general feeling that the community gives. It feel like I am part of a big people- powered movement. It feels like a revolution.*
>
> *I felt a strong moral and ethical imperative that technology should serve people.*
>
> *Help others to have healthier life.*

Comorbidities

> *I started on AndroidAPS when I was diagnosed with cancer needing chemotherapy. I found it extremely beneficial especially for those times when I was at my lowest and unable to control my BGs in the old way because of insulin resistance. Also when I was admitted to hospital because of infections and sepsis it was a godsend.*
>
> *More beneficial sexual activity, PDE-5 inhibitors no longer required.*
>
> *Because of other conditions, I have to take cortisone in different doses on a regular basis. This has made my diabetes management so difficult. The loop absorbs my BG fluctuations much better.*
>
> *Achalasia (food gets stuck in the esophagus at night), making blood sugar uncontrollable.*
>
> *I have been on a pump since 1992. I was on the 670G for over a year, and I felt helpless in my efforts to achieve excellent glycemic control while still living my random and not standardized life, where I eat when I am hungry, or forget to eat, and where pre-bolusing is dangerous, because I also have ADHD and I have forgotten to eat many times. My insulin needs vary depending on what I do in terms of activity, but also randomly on the day of the week, the time of the month and many other factors that i don't understand. On the 670G every weekend of high physical activity was followed by a couple of days of high BGs due to the user's inability to interact with the proprietary algorithms (Oh I am so done with Medtronic now).*

Diabetes-Related Complications

> *After 29 years of MDI and [...] retinopathy I decided to improve my health. I've researched several ways to improve control. Ultimately autonomy is the box I needed ticked! AAPS ticks that box 100%.*

> *Gastroparesis, I barely had nights where I wasn't over 200 half the night. With the G5 I was woken up at 170 and was able to intervene. Since the loop and some completed goals, I fall asleep again because the loop prevents the uncontrollable rise!*
>
> *Heart operation after 30+ years of poorly controlled diabetes.*

Female Health

> *Wanted better control for pregnancy.*
>
> *As someone whose hormone levels are not considered standard and rapidly change, the ability to [have] a helping hand to smooth out these Diabetes related complications (notably hyperglycemia episodes) was very important to me, as the situation is never the same twice and requires different treatment on a day-to-day basis.*
>
> *Deteriorating* HbA_{1c} *due to puberty and insulin resistance. Massive amounts of insulin needed giving unpredictable blood glucose.*
>
> *After manifestation of T1D, we made very high demands on* HbA_{1c} *and TiR for the benefit of our daughter...but with the onset of puberty, this led to an almost impossible workload (correcting 10-15 times at night).*
>
> *To have more insight as to why my blood glucose was so volatile due to changing hormones (menopause).*

Out-of-Pocket Expenses

> *It was questionable whether I would meet the health insurer's criteria for the Minimed 670 system for reimbursement. I don't have a CGM either, just the Freestyle Libre with an additional transmitter.*

Improving Performance

> *To improve my work at the office.*
>
> *To improve athletic performance by controlling night time blood sugars.*

Curiosity

> *To learn more about my diabetes in general. You have to acquire a lot of knowledge (technical as well as physiological aspects) before you start looping, and you get excellent support from developers and the community.*
>
> *The fact alone that you can be curious again about something new to the diabetes field, to see a form of therapy as an exciting challenge, plus the (so far not yet fulfilled hope) to finally better control the hardly controllable variating [postprandial] values.*

Most of the indicated *other motivations* provided greater details about the 14 predefined statements. The most frequently mentioned motivations for all—adults and caregivers—were *better management* and *reducing the disease burden*. The first motivation appears consistent with several statements related to hypoglycemia and hyperglycemia and risk reduction, whereas the second motivation may correspond with *to put diabetes management more on auto-pilot and interact less frequently with the therapy system*. This aspect and sleep quality are understood as the quality of life gains. Of motivations not covered by the predefined responses, the most frequently mentioned was *autonomy gain* in both adults and children or adolescents, as indicated by the caregivers. All these aspects were associated with improvements in family life:

> *This is for my wife. She wants me to live forever, this is as close as I can do for her.*

Psychosocial aspects, ranging from diabetes burnout and distress to a desire to improve athletic performance to increasing efficacy at work, were also identified as important motivating factors. The following comments illustrate the wide-ranging benefits experienced by many participants after adopting the technology:

> *Management of diabetes is helped by support, but it is very much a self-managed disease and requires 24/7 attention. Closed looping makes it just so much better, much of the time I can leave [AndroidAPS] to take care of basals by itself. Quality of life is so much better. I can sleep without worrying about not waking up because of a bad hypo. [...] I am so grateful to all the software developers who have freely given their expertise and time to make this possible.*

As highlighted in the example above, an important role was played by the *community spirit* and peer support in social networks:

> *Something that also influenced me [...] was the support from the community, and the general feeling that the community gives. It feels like I am part of a big people-powered movement. It feels like a revolution.*

Not only a *Do-It-Yourself* mindset and being *early adopters* of technology but also being motivated and empowered to improve one's life were frequently mentioned:

> *Because it's the most natural thing to do, after getting to know that it's possible. Because I could.*

Some motivations included other health-related aspects such as improving the management of existing diabetes-related complications and increasing safety by avoiding severe hypoglycemia. Other comorbidities, such as cancer, sexual health difficulties, or conditions requiring cortisone treatment, were also mentioned:

I have to take cortisone in different doses on a regular basis. This has made my diabetes management so difficult. The loop absorbs my BG fluctuations much better.

Women and caregivers of female children highlighted female health aspects such as hormone-related changes in insulin sensitivity, family planning, and pregnancy as reasons to commence open-source AID:

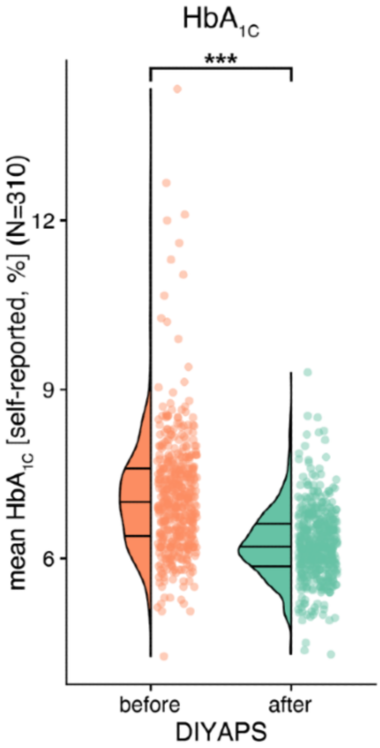
> *To have more insight as to why my BG was so volatile due to changing hormones (menopause).*

For some, special features were only offered by open-source AID and not by commercial systems, that is, customizable targets and the option to bolus from a smartwatch. For caregivers, remote real-time access to their child's data and the option to remotely control their child's AID system have been frequently described.

Improved Glycemic Outcomes Across All Age Groups and Genders

To assess glycemic outcomes, participants were asked to report their or their child's 3 most recent HbA_{1c} results before as well as the first 3 HbA_{1c} results after commencing the open-source AID. HbA_{1c} levels decreased significantly following open-source AID implementation ($P<.001$) from an average of 7.14% (SD 1.13%; mean 54.5 mmol/mol, SD 12.4) to 6.24% (SD 0.64%; mean 44.7 mmol/mol, SD 7.0), with an effect size of −0.9% (Figure 2).

Figure 2. Positive effects of open-source automated insulin delivery on clinical outcomes: average self-reported glycated hemoglobin (%; y-axis) for all 310 respondents, before and after open-source automated insulin delivery (x-axis) distinguished by orange and green colors, respectively. The left side is displayed as a density plot, with horizontal lines indicating quartiles. The right side depicts the data as a scatter plot. DIYAPS: Do-it-Yourself Artificial Pancreas System; HbA_{1c}: glycated hemoglobin.



The average self-reported TIR across adults and children with diabetes significantly increased by +17.4%, from 62.96% (SD 16.18%) to 80.34% (SD 9.41%; $P<.001$; Figure 3). Similar outcomes were observed separately for adults and children with diabetes (Multimedia Appendix 3) and were independent of age and sex (Figure 4). Overall, 92.3% (286/310) of the respondents reported a decreased average HbA_{1c} level (Figure 5).

Figure 3. Self-reported time in range (%; x-axis) of adults and children with diabetes, before and after implementing an open-source automated insulin delivery system. The left side is displayed as a density plot, with horizontal lines indicating quartiles. The right side depicts the data as a scatter plot.

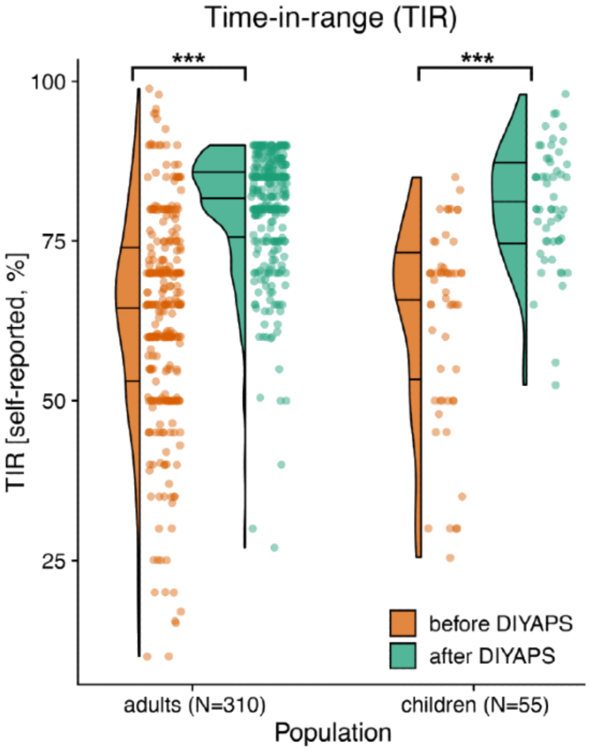


Figure 4. Improvements in self-reported glycated hemoglobin levels associated with open-source automated insulin delivery, independent of age or gender: relation between average glycated hemoglobin levels (%; y-axis) and age (x-axis), shown separately for female and male respondents (top and bottom rows, respectively). Colors separate average glycated levels before (orange) and after (green) open-source automated insulin delivery implementation. Each point represents one respondent after filtering of responses (the Methods section). Solid lines and their gray areas represent the trend and standard error for the respective groups. AID: automated insulin delivery; HbA_{1c}: glycated hemoglobin.

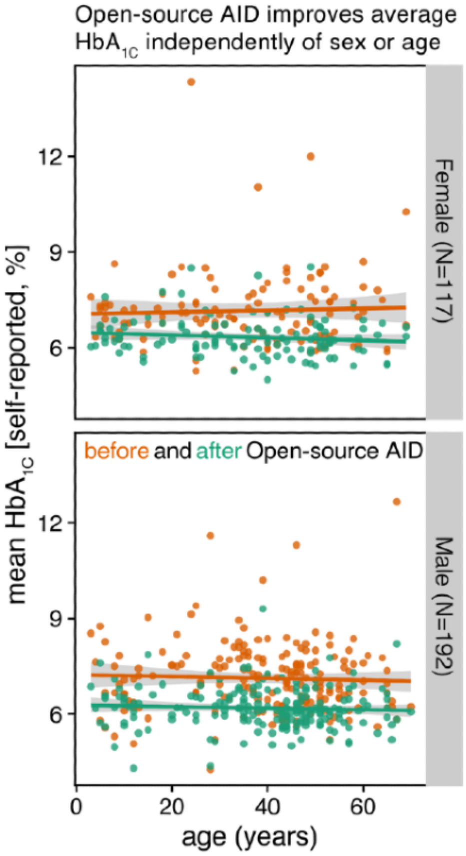
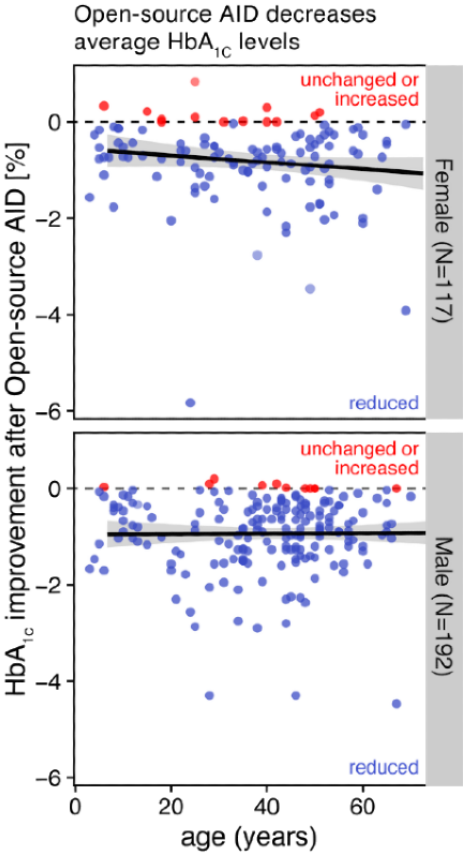


Figure 5. Improvements in self-reported glycated hemoglobin levels associated with open-source automated insulin delivery, independent of age or gender: the y-axis shows the difference of average glycated hemoglobin levels after open-source automated insulin delivery, compared with before its implementation. Colors distinguish respondents with reduced average glycated hemoglobin (blue) from those with unchanged or increased glycated hemoglobin (red). AID: automated insulin delivery; HbA_{1c}: glycated hemoglobin.



Discussion

Principal Findings

This study is the first to systematically analyze the motivations found within the #WeAreNotWaiting movement of people with diabetes, who have built and maintained their open-source AID systems and created their own ecosystem of international self-support networks. To the best of the authors' knowledge, this is also the largest study reporting the self-reported clinical outcomes of open-source AID users across several continents. We found large effect sizes for self-reported improvements in HbA_{1c} (−0.9% on average) and TIR (+17.4% on average), indicating considerable biomedical benefits associated with open-source AID, which were independent of sex and age.

Why #WeAreNotWaiting: Main Motivators to Choose an Open-source AID

The main motivators for adults were improvements in overall glycemic and long-term outcomes and quality of life, whereas the strongest motivation for caregivers was improvement of their own sleep, followed by improved glycemia of the child and possibility of remotely controlling glycemia and insulin delivery via the internet. The results indicate that motivations are configured differently among caregivers and that other motivations also scored a high level of consensus among the respondents. These findings suggest that motivation to transition toward open-source AID is multifaceted and complex, with reasoning and decision making bound up with the psychological and social intricacies of individuals' lives.

Improvement in Sleep Quality

Caregivers experience reduced sleep quality because of fear of hypoglycemia, which often requires them to regularly check their child's glucose levels overnight [22]. In our study, caregivers reported experiencing fewer demands and less apprehensiveness regarding their child's glucose levels at nighttime. As shown in the free-text responses, open-source AID also appears to offer caregivers with the reassurance necessary to provide their child more autonomy and engage in activities that might otherwise present a risk, such as having *a sleepover* with friends. Previous studies in adults using open-source AID have shown self-reported improvement in sleep quality [23,24]. These initial findings indicate a substantial benefit for users and caregivers for sleep and most likely for their psychological and physical well-being. Poor quality of sleep negatively affects the psychological well-being, cognitive functioning, and a diverse range of hormones that affect the regulation of appetite and our homeostatic systems as well as the immune system [25-34]. Recent research also points to sleep as impacting the actual maintenance of the brain and our DNA regenerative systems [35]. Thus, AID may play an important role in improving the psychological and physical health of people with diabetes and their family members. However, it has been noted elsewhere that the potential discomfort or inconvenience of wearing devices and overnight alarms may also hamper the benefits for some users [36].

The Importance of Customizability: One Size Does Not Fit All

The majority of participants reported that currently approved and available commercial therapy options may not be sufficiently flexible or customizable to fulfill their or their children's individual needs. Among caregivers, features only available in open-source AID, in particular, the possibility of remote management was the main additional motivation. A wider range of features and adjustable settings to improve user experience may be beneficial for people with diabetes of all ages, which mirrors a recent study in very young children using a commercial AID [36]. Interestingly, for many adult respondents, *curiosity* was cited as an important motivation. In contrast, curiosity was a much lower motivating factor for caregivers who chose to build a system for more practical or psychosocial reasons.

Do-it-Yourself Is Not Do-it-Alone: The Impact of Peer Support

The ability to receive and provide support within the do-it-yourself (DIY) community and observe the success of others was an important motivating factor associated with opting to use open-source AID for some. Obtaining and exchanging information and advice from open forums limits the spread of misinformation because other users constitute a community of inquirers ready to challenge and correct spurious or misleading information [37]. Although open-source AID is individualized and patient focused, it is also a grassroots community-driven movement. The number of responses to our survey reflects the enthusiasm and importance of open-source AID. In challenging traditional top-down hierarchies in medicine, open-source AID presents a patient-focused initiative that serves to empower people with diabetes through personalized technology. Because of the availability of current technology and individualized innovations, open-source AID has previously been described as having the potential to democratize health care, revolutionizing treatment and the way people with diabetes as well as other stakeholders such as care teams, researchers, and device manufacturers view chronic conditions such as diabetes [38]. The importance of peer support in the context of open-source AID use has recently been highlighted elsewhere, and a sense of community underpinning the development and diffusion of open-source AID has been emphasized by individual users [39]. Further research should consider community phenomena as an integral part of the DIY experience.

Strengths and Limitations

This study is the first to investigate motivations of users and caregivers to build and use open-source AID. In addition, this is the largest study that reports self-reported clinical outcomes of open-source AID users globally and adds to the existing evidence base around glycemic outcomes in smaller cohorts [11-15]. At the time of data collection, it surveyed the majority of open-source AID users worldwide, with 897 respondents of a population estimated in 2018 (N=1500). The sample is impressive not only in size but also because people with diabetes from various continents and regions of the world are represented. Of other strengths, this study has been conducted by an interdisciplinary consortium, with members of the diabetes community directly involved. However, this firsthand experience should be acknowledged as a potential bias. In addition, a key limitation of the study is the fact that self-reported outcomes have not been corroborated by clinical records. Some may consider this has potential for inaccuracy, that is, by lacking precision as witnessed by the overaccumulation of rounded TIR values or to be biased by the specificity of the population that participated. However, other studies have found that real-world data are robust and reliable [40]. We acknowledge that open-source AID users constitute a specific group of people with diabetes who may be highly motivated, engaged, and willing to improve the quality of their diabetes care and life, limiting the scope of our findings to this group. However, a recent study of newly available commercial AID systems indicates that users are similarly motivated to achieve the best possible outcomes but are dissatisfied with postprandial glycemic outcomes under commercial AID systems [41]. The earliest adopters of available commercial AID technology may be more similar to the group of people with diabetes choosing open-source AID in terms of motivation and engagement than expected. The lack of widespread availability of AID technology in general, including commercial systems with regulatory approval, at the time of the study likely influenced the perspectives of people with diabetes choosing open-source systems. In the future, wider availability—and, importantly, better funding or insurance coverage of commercial AID systems—may further influence this cohort. Similarly, it remains to be seen if the predicted *second generation* of commercial systems—with a hypothesis of increased sophistication or improvements on the first-generation devices and algorithms—will enable people with diabetes to achieve results similar to those they are currently achieving with their chosen open-source AID system.

It should also be noted that those who benefit from and continue to use open-source AID may be motivated to share their positive experiences. Although the survey was open to people with type 2 and gestational diabetes, it was completed almost exclusively by adults and caregivers of children with type 1 diabetes. This is likely a reasonable reflection of the DIY community, but efforts need to be made in the future to encourage participation of those with other types of diabetes. The high percentage of respondents from Europe may be influenced by the fact that the majority of the research team is EU based, which may be another bias. This may also be explained by the characteristics of the European health services provision and reimbursement of diabetes-related technology, which may provide a greater degree of accessibility of the underlying components needed (eg, pumps and continuous glucose monitors). Language barriers may have limited responses from other parts of the world as the survey was only available in 2 languages. Finally, the majority of participants had a university degree, suggesting that open-source AID uptake is more common among people of higher socioeconomic status. Increasing socioeconomic inequalities in access to the underlying technologies needed to build an open-source AID system may help to explain some of these variations. Thus, further investigation into how the wider diffusion of open-source AID is conditioned by factors such as social class, gender, age, and geographic location is required.

Conclusions

This study provides new insights into the factors that motivate people to adopt *DIY* solutions in relation to their diabetes and beyond. Our findings contribute to a better understanding of the unmet needs of people with diabetes and some of the current challenges in the uptake of AID technology. This study, alongside other efforts in the DIY community space, can help key stakeholders, including academia, the medical device industry, regulators, health care providers, and care teams, to better address some of the fundamental gaps and needs that still exist for people with diabetes worldwide, even with the advent of first-generation commercial AID systems. The DIY movement has resulted in impactful solutions addressing the unmet needs of people with diabetes and represents an exemplary case of how informed and connected patients are shaping the direction of technological innovation in diabetes care and potentially for other areas of health care in the future.

Acknowledgments

The authors would like to thank the #WeAreNotWaiting and DIY community that supported and greatly contributed to this study. The OPEN project has received funding from the European Commission’s Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie Action Research and Innovation Staff Exchange grant agreement number 823902. J Speight and CH were supported by the core funding to the Australian Center for Behavioral Research in Diabetes provided by the collaboration between Diabetes Victoria and Deakin University. KB received the funding from the Berlin Institute of Health Digital Clinician Scientist Program and the SPOKES Wellcome Trust Translational Partnership Program. The funding sources had no involvement in the study.

Authors' Contributions

KB was the guarantor of this study. KB, DL, TS, HL, KR, J Schipp, KAG, CH, and BC performed the literature search. KB, SO, BC, A Tappe, DL, and KR designed the survey. KB collected the data. KB, DL, SO, TF, KAG, and BC analyzed and interpreted the data. A Thieffry conducted quantitative analyses and produced the associated figures. KB, DL, SO, and BC prepared the initial draft. All authors reviewed the manuscript, contributed revisions, and approved the final version.

Conflicts of Interest

All authors have completed the Unified Competing Interest form and declare the following: KB has received research grants from the Berlin Institute of Health, SPOKES Wellcome Trust, Stiftung Charité, and the German Diabetes Association (Deutsche Diabetes Gesellschaft). She has served as a speaker and advisory board member for Medtronic Diabetes and Hi.Health and received fees for medical consulting and public speaking from Roche Diabetes Care, Dexcom, Medtronic Diabetes, Diabeloop, BCG Digital Ventures, and Bertelsmann Stiftung, outside the submitted work. DL reports grants from the Robert Wood Johnson Foundation, JDRF, personal fees from Lilly, Diabeloop, Roche Diabetes Care, and Novo Nordisk and Tandem, outside the submitted work. J Speight has served on advisory boards for Janssen, Medtronic, Roche Diabetes Care, and Sanofi Diabetes; her research group (Australian Centre for Behavioural Research in Diabetes) has received honoraria for this advisory board participation and has also received unrestricted educational grants and in-kind support from Abbott Diabetes Care, AstraZeneca, Medtronic, Roche Diabetes Care, and Sanofi Diabetes. J Speight has also received sponsorship to attend educational meetings from Medtronic, Roche Diabetes Care, and Sanofi Diabetes and consultancy income or speaker fees from Abbott Diabetes Care, AstraZeneca, Medtronic, Novo Nordisk, Roche Diabetes Care, and Sanofi Diabetes, all outside the submitted work. KR is an advisory board member of Lilly Diabetes Care and Abbott Diabetes Care outside the submitted work. KB, DL, SO, A Tappe, and KAG are members of the open-source AID web-based community. All other authors have no conflict of interest to declare.

Multimedia Appendix 1
Questionnaire for people with diabetes.
[PDF File (Adobe PDF File), 92 KB - jmir_v23i6e25409_app1.pdf]

Multimedia Appendix 2
Questionnaire for caregivers of children and adolescents with diabetes.
[PDF File (Adobe PDF File), 92 KB - jmir_v23i6e25409_app2.pdf]

Multimedia Appendix 3
Supplementary tables 1 (codes and frequency of mentioned additional motivations as responses to the open-ended question) and 2 (detail of time-in-range improvements following open-source AID implementation).
[DOCX File , 16 KB - jmir_v23i6e25409_app3.docx]

References

1. The Diabetes Control and Complications Trial Research Group. Influence of intensive diabetes treatment on quality-of-life outcomes in the diabetes control and complications trial. Diabetes Care 1996 Mar;19(3):195-203. [doi: 10.2337/diacare.19.3.195] [Medline: 8742561]
2. Miller RG, Secrest AM, Sharma RK, Songer TJ, Orchard TJ. Improvements in the life expectancy of type 1 diabetes: the Pittsburgh Epidemiology of Diabetes Complications study cohort. Diabetes 2012 Nov;61(11):2987-2992 [FREE Full text] [doi: 10.2337/db11-1625] [Medline: 22851572]
3. Jørgensen ME, Almdal TP, Carstensen B. Time trends in mortality rates in type 1 diabetes from 2002 to 2011. Diabetologia 2013 Nov;56(11):2401-2404. [doi: 10.1007/s00125-013-3025-7] [Medline: 23949580]
4. 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]
5. American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes-2018. Diabetes Care 2018 Jan;41(Suppl 1):55-64. [doi: 10.2337/dc18-S006] [Medline: 29222377]
6. Bjerg L, Hulman A, Carstensen B, Charles M, Jørgensen ME, Witte DR. Development of microvascular complications and effect of concurrent risk factors in type 1 diabetes: a multistate model from an observational clinical cohort study. Diabetes Care 2018 Nov;41(11):2297-2305. [doi: 10.2337/dc18-0679] [Medline: 30131399]
7. Polonsky WH, Anderson BJ, Lohrer PA, Welch G, Jacobson AM, Aponte JE, et al. Assessment of diabetes-related distress. Diabetes Care 1995 Jun;18(6):754-760. [doi: 10.2337/diacare.18.6.754] [Medline: 7555499]
8. Hagger V, Hendrieckx C, Sturt J, Skinner TC, Speight J. Diabetes distress among adolescents with type 1 diabetes: a systematic review. Curr Diab Rep 2016 Jan;16(1):9. [doi: 10.1007/s11892-015-0694-2] [Medline: 26748793]
9. Buchberger B, Huppertz H, Krabbe L, Lux B, Mattivi JT, Siafarikas A. Symptoms of depression and anxiety in youth with type 1 diabetes: a systematic review and meta-analysis. Psychoneuroendocrinology 2016 Aug;70:70-84. [doi: 10.1016/j.psyneuen.2016.04.019] [Medline: 27179232]
10. Skinner TC, Joensen L, Parkin T. Twenty-five years of diabetes distress research. Diabet Med 2020 Mar;37(3):393-400. [doi: 10.1111/dme.14157] [Medline: 31638279]
11. Lewis D, Leibrand S, #OpenAPS Community. Real-world use of open source artificial pancreas systems. J Diabetes Sci Technol 2016 Nov;10(6):1411 [FREE Full text] [doi: 10.1177/1932296816665635] [Medline: 27510442]
12. 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;21(10):2333-2337. [doi: 10.1111/dom.13810] [Medline: 31183929]
13. Petruzelkova L, Soupal J, Plasova V, Jiranova P, Neuman V, Plachy L, et al. Excellent glycemic control maintained by open-source hybrid closed-loop AndroidAPS during and after sustained physical activity. Diabetes Technol Ther 2018 Nov;20(11):744-750. [doi: 10.1089/dia.2018.0214] [Medline: 30285476]
14. 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]
15. 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]
16. Toffanin C, Kozak M, Sumnik Z, Cobelli C, Petruzelkova L. In silico trials of an open-source android-based artificial pancreas: a new paradigm to test safety and efficacy of Do-It-Yourself systems. Diabetes Technol Ther 2020 Feb;22(2):112-120. [doi: 10.1089/dia.2019.0375] [Medline: 31769699]
17. O'Donnell S, Lewis D, Fernández MM, 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 [FREE Full text] [doi: 10.2196/15368] [Medline: 31742563]
18. 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]
19. 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]
20. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
21. 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]
22. Sinisterra M, Hamburger S, Tully C, Hamburger E, Jaser S, Streisand R. Young children with type 1 diabetes: sleep, health-related quality of life, and continuous glucose monitor use. Diabetes Technol Ther 2020 Aug;22(8):639-642. [doi: 10.1089/dia.2019.0437] [Medline: 32027177]
23. Litchman ML, Lewis D, Kelly LA, Gee PM. Twitter analysis of #OpenAPS DIY artificial pancreas technology use suggests improved A1C and quality of life. J Diabetes Sci Technol 2019 Mar;13(2):164-170 [FREE Full text] [doi: 10.1177/1932296818795705] [Medline: 30198751]

24. Palmer W, Greeley SA, Letourneau-Freiberg LR, Naylor RN. Using a Do-It-Yourself artificial pancreas: perspectives from aatients and diabetes providers. J Diabetes Sci Technol 2020 Sep;14(5):860-867. [doi: 10.1177/1932296820942258] [Medline: 32680447]
25. Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 2003 Mar 15;26(2):117-126. [doi: 10.1093/sleep/26.2.117] [Medline: 12683469]
26. Spiegel K, Sheridan JF, Van Cauter E. Effect of sleep deprivation on response to immunization. J Am Med Assoc 2002 Sep 25;288(12):1471-1472. [doi: 10.1001/jama.288.12.1471-a] [Medline: 12243633]
27. Besedovsky L, Lange T, Haack M. The sleep-immune crosstalk in health and disease. Physiol Rev 2019 Jul 01;99(3):1325-1380 [FREE Full text] [doi: 10.1152/physrev.00010.2018] [Medline: 30920354]
28. Dørheim SK, Bondevik GT, Eberhard-Gran M, Bjorvatn B. Sleep and depression in postpartum women: a population-based study. Sleep 2009 Jul;32(7):847-855 [FREE Full text] [doi: 10.1093/sleep/32.7.847] [Medline: 19639747]
29. Prinz P. Sleep, appetite, and obesity--what is the link? PLoS Med 2004 Dec;1(3):e61 [FREE Full text] [doi: 10.1371/journal.pmed.0010061] [Medline: 15599390]
30. Gottlieb DJ, Punjabi NM, Newman AB, Resnick HE, Redline S, Baldwin CM, et al. Association of sleep time with diabetes mellitus and impaired glucose tolerance. Arch Intern Med 2005 Apr 25;165(8):863-867. [doi: 10.1001/archinte.165.8.863] [Medline: 15851636]
31. Zhai L, Zhang H, Zhang D. Sleep duration and depression among adults: a meta-analysis of prospective studies. Depress Anxiety 2015 Sep;32(9):664-670. [doi: 10.1002/da.22386] [Medline: 26047492]
32. Lim J, Dinges DF. A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. Psychol Bull 2010 May;136(3):375-389 [FREE Full text] [doi: 10.1037/a0018883] [Medline: 20438143]
33. Lin J, Jiang Y, Wang G, Meng M, Zhu Q, Mei H, et al. Associations of short sleep duration with appetite-regulating hormones and adipokines: a systematic review and meta-analysis. Obes Rev 2020 Nov;21(11):e13051. [doi: 10.1111/obr.13051] [Medline: 32537891]
34. Irwin MR, Olmstead R, Carroll JE. Sleep disturbance, sleep duration, and inflammation: a systematic review and meta-analysis of cohort studies and experimental sleep deprivation. Biol Psychiatry 2016 Jul 01;80(1):40-52 [FREE Full text] [doi: 10.1016/j.biopsych.2015.05.014] [Medline: 26140821]
35. Jessen NA, Munk AS, Lundgaard I, Nedergaard M. The glymphatic system: a beginner's guide. Neurochem Res 2015 Dec;40(12):2583-2599 [FREE Full text] [doi: 10.1007/s11064-015-1581-6] [Medline: 25947369]
36. 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;20(6):794-799 [FREE Full text] [doi: 10.1111/pedi.12872] [Medline: 31140654]
37. White K, Gebremariam A, Lewis D, Nordgren W, Wedding J, Pasek J, et al. Motivations for participation in an online social media community for diabetes. J Diabetes Sci Technol 2018 May;12(3):712-718 [FREE Full text] [doi: 10.1177/1932296817749611] [Medline: 29310448]
38. Burnside M, Crocket H, Mayo M, Pickering J, Tappe A, de Bock M. Do-It-Yourself automated insulin delivery: a leading example of the democratization of medicine. J Diabetes Sci Technol 2020 Sep;14(5):878-882 [FREE Full text] [doi: 10.1177/1932296819890623] [Medline: 31876179]
39. Tenderich A, Tenderich B, Barton T, Richards SE. What Are PWDs (People With Diabetes) doing online? A netnographic analysis. J Diabetes Sci Technol 2019 Mar;13(2):187-197 [FREE Full text] [doi: 10.1177/1932296818813192] [Medline: 30477335]
40. Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. Value Health 2007;10(5):326-335 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00186.x] [Medline: 17888097]
41. Lewis D. Automated Insulin Delivery in Real Life (AID-IRL): real-world user perspectives on commercial AID. J Diabetes Sci Technol 2020 Sep 16:1932296820957360. [doi: 10.1177/1932296820957360] [Medline: 32935554]

Abbreviations

AID: automated insulin delivery
DIY: do-it-yourself
DIYAPS: Do-It-Yourself Artificial Pancreas Systems
HbA$_{1c}$: glycated hemoglobin
OPEN: Outcomes of Patients’ Evidence with Novel, Do-it-Yourself Artificial Pancreas Technology
TIR: time in range

Edited by R Kukafka; submitted 31.10.20; peer-reviewed by S Hussain, R Lal, A Garrity, J Paul; comments to author 10.12.20; revised version received 19.12.20; accepted 16.03.21; published 07.06.21.

Please cite as:
Braune K, Gajewska KA, Thieffry A, Lewis DM, Froment T, O'Donnell S, Speight J, Hendrieckx C, Schipp J, Skinner T, Langstrup H, Tappe A, Raile K, Cleal B
Why #WeAreNotWaiting—Motivations and Self-Reported Outcomes Among Users of Open-source Automated Insulin Delivery Systems: Multinational Survey
J Med Internet Res 2021;23(6):e25409
URL: https://www.jmir.org/2021/6/e25409
doi:10.2196/25409
PMID:34096874

Original Paper

Medical Insights from Posts About Irritable Bowel Syndrome by Adolescent Patients and Their Parents: Topic Modeling and Social Network Analysis

Bu Zhong[1], PhD; Qian Liu[2], BEng, MSc, PhD

[1]Donald P. Bellisario College of Communications, Pennsylvania State University, University Park, PA, United States

[2]School of Journalism and Communication, National Media Experimental Teaching Demonstration Center, Jinan University, Guangzhou, Guangdong Province, China

Corresponding Author:
Qian Liu, BEng, MSc, PhD
School of Journalism and Communication
National Media Experimental Teaching Demonstration Center
Jinan University
601 Huangpu West Ave, Tianhe District
Guangzhou, Guangdong Province, 510632
China
Phone: 86 13302292599
Email: tsusanliu@jnu.edu.cn

Abstract

Background: Adolescents with irritable bowel syndrome (IBS) are increasingly seeking and sharing information about their symptoms in web-based health care forums. Their posts and those from their parents contain critical insights that can be used by patients, physicians, and caregivers to manage IBS symptoms.

Objective: The aim of this study is to examine the posts from adolescent patients and their parents in a health forum, IBS Group, to better understand the key challenges, concerns, and issues of interest to young patients with IBS and their caregivers.

Methods: Using topic modeling and social network analysis, in this study, we analyzed all the messages (over 750 topics and 3400 replies) posted on the IBS Group forum from 2010-2019 by adolescents with IBS aged 13-17 years and parents having children with IBS. We first detected 6 major topics in the posts by adolescent patients and parents on teenagers' IBS symptoms and the interaction between the topics. Social network analysis was then performed to gain insights into the nature of web-based interaction patterns among patients and caregivers.

Results: Using the Latent Dirichlet Allocation algorithm and a latent Dirichlet allocation visualization tool, this study revealed 6 leading topics of concern in adolescents with IBS: school life, treatment or diet, symptoms, boys' ties to doctors, social or friend issues, and girls' ties to doctors. The top 6 topics in the parents' discussions were school life, girls' issues, boys' issues, diet choice, symptoms, and stress. The analyses show that the adolescent patients themselves are most concerned about the effect of IBS on their everyday activities and social lives. For parents having daughters with IBS, their top concerns were related to the girls' school performance and how much help they received at school. For their sons, the parents were more concerned about the pain and suffering that their sons had to endure. Both parents and adolescents gained social support from the web-based platform. Topic modeling shows that IBS affects teenagers the most in the areas of pain and school life. Furthermore, the issues raised by parents suggest that girls are bothered more by school performance over pain, whereas boys show exactly the opposite: pain is of greater concern than school performance.

Conclusions: This study represents the first attempt to leverage both machine learning approaches and social network analysis to identify top IBS concerns from the perspectives of adolescent patients and caregivers in the same health forum. Young patients with IBS must face the challenges of social influences and anxiety associated with this health disorder in addition to physical pain and other symptoms. Boys and girls are affected differently by pain and school performance and view the IBS impacts differently from the parents.

(J Med Internet Res 2021;23(6):e26867) doi:10.2196/26867

KEYWORDS

irritable bowel syndrome; health care forum; adolescent; parents; topic modeling; social network analysis

Introduction

Background

More than 58 million Americans are affected by irritable bowel syndrome (IBS) [1], a gastrointestinal disorder that causes abdominal pain, bloating, gas, and diarrhea or constipation. However, IBS in adolescents is a less known problem than it is among adults, although 6% of middle school students and 14% of high school students in the United States have symptoms indicative of IBS [2]. Patients with IBS, similar to others with chronic medical conditions, are increasingly seeking and sharing health information through web-based health care forums to manage their symptoms [3]. The forums, functioning similarly to online support groups, have served as a critical source of health information for patients and caregivers. On the internet, patients with gastrointestinal disorders could seek social, emotional, and peer support in addition to the real-time exchange of information related to medical concerns and symptom management [4].

Using data from the IBS Group forum, prior studies have investigated how the ways in which patients with IBS sought out or offered social support affected their behavior toward others on the forum [3,5]. For instance, expressing one's emotions in a post during support seeking significantly increases the likelihood of receiving social support from others who are active in the web-based forum, but reciprocating support with a fixed set of peers discourages individuals from continued social support to others in the larger community [3]. User participation in the forum discussion was found to follow a power law, meaning that a small number of users on the forum became opinion leaders and generated most of the activity on the forum.

IBS in Adolescents

IBS symptoms can be destructive to both adolescents and adults. Despite the serious impact of IBS on patients' quality of life, this gastrointestinal disorder remains a clinical challenge that lacks diagnostic, radiologic, or laboratory markers [3]. To continue this stream of research, this research proposes to use probabilistic topic models for text summarization and topic detection by analyzing an unlabeled corpus of adolescent patients' posts on their IBS symptoms and their parents' posts on how their children coped with IBS. Specifically, we use the latent Dirichlet allocation (LDA) approach, a generative unsupervised probabilistic algorithm that isolates the top K topics in a data set, as described by the most relevant N keywords [6]. Thus, the posts and replies to them in the corpora are treated as random mixtures of latent topics, where each topic is characterized by a latent Dirichlet distribution over a fixed vocabulary. Here, the word *latent* means that topics must be inferred from keywords rather than directly observed.

Using the LDA topic modeling method [7], in this research, we first detected major topics in parents' posts on the IBS Group forum discussing the IBS symptoms experienced by their teenaged sons and daughters. All the topics uncovered in the LDA analyses were analyzed and visualized to explore the relationships among the topics, whose results would assist in medical decision making for the teenagers, their parents, and health professionals. A social network analysis was also performed to gain insights into the nature of web-based interaction patterns.

The remainder of this paper is organized as follows. The next section provides the research background by reviewing the related literature. In the following section, we discuss the research questions (RQs) and explain the methodology. Subsequent sections present the main results and conduct a social network analysis. On the basis of the analyses, the findings and their implications, including the limitations, are examined in the *Discussion* section. Interventions have been proposed to mitigate IBS symptoms in adolescents. Finally, the conclusions are presented in the final section.

IBS Diagnosis

IBS diagnosis is currently based on a symptom complex, and there is no definitive diagnostic test. In children and adolescents, the prevalence of IBS-associated abdominal pain ranges from 6% to 17% and 13% to 38%, respectively [2,8-10]. Previous studies of information flow on public patient-centered health care forums—namely, forums initiated and maintained by patients—have suggested that patients with IBS rely on health information on such forums to evaluate their medical decisions, health risks, and treatment options as well as seeking emotional support [11].

Children and adolescents are particularly vulnerable to IBS. Compared with adult patients, children and adolescents with IBS are distinct in the prevalence of IBS subtypes [12] and are typically prescribed with different pharmacologic and dietary management plans than adult patients [13]. Hence, physicians have difficulty in confirming their diagnoses [14] and may overlook some of the associations among multiple symptoms and facets of this health problem. The limited long-term treatment options deemed safe for younger patients, coupled with the intermittent and unpredictable recurrence of symptoms, result in a chronic state of uncertainty that has profound and formative impact on children's and adolescents' school performance and their quality of life, particularly in eroding younger patients' self-esteem and magnifying their insecurity as *misfits* in schools and society. Although a large amount of work has been devoted to understanding IBS issues in adult patients, challenges in managing pediatric IBS symptoms are only beginning to garner the attention of researchers and health care practitioners.

Previous studies on adolescents' IBS issues are helpful in understanding the pathophysiology of the disease and the etiology of abdominal functional pain, food-induced gastrointestinal symptoms, and other dietary consequences [2]. However, with rare exceptions [15], most studies did not consider possible complications that arise from the symptoms and everyday management of IBS during vulnerable developmental phases, such as childhood and adolescence. The novel aspects of our work are that it is the first study to leverage both machine learning approaches (eg, LDA [7,16]) and social network analysis to identify top IBS concerns from the perspectives of children, adolescents, and caregivers on web-based health forums. Such state-of-the-art approaches should be able to provide novel data-driven insights into factors that facilitate and impede the translation and uptake of health care knowledge and practices.

Therefore, this study proposed the following RQs to explore informational insights by analyzing the posts of adolescents with IBS and their parents who discussed the impact of IBS on young patients:

- RQ1. What are the top 6 topics in parents' posts regarding the impact of IBS on their children?
- RQ2. What are the top 6 topics in teenagers' posts on IBS symptoms?
- RQ3. Are there any sex differences between the concerns that parents post on the forum about their male and female children?
- RQ4. What is the nature of the interactions in the social network on the IBS health forum and what insights can we learn from it?

Methods

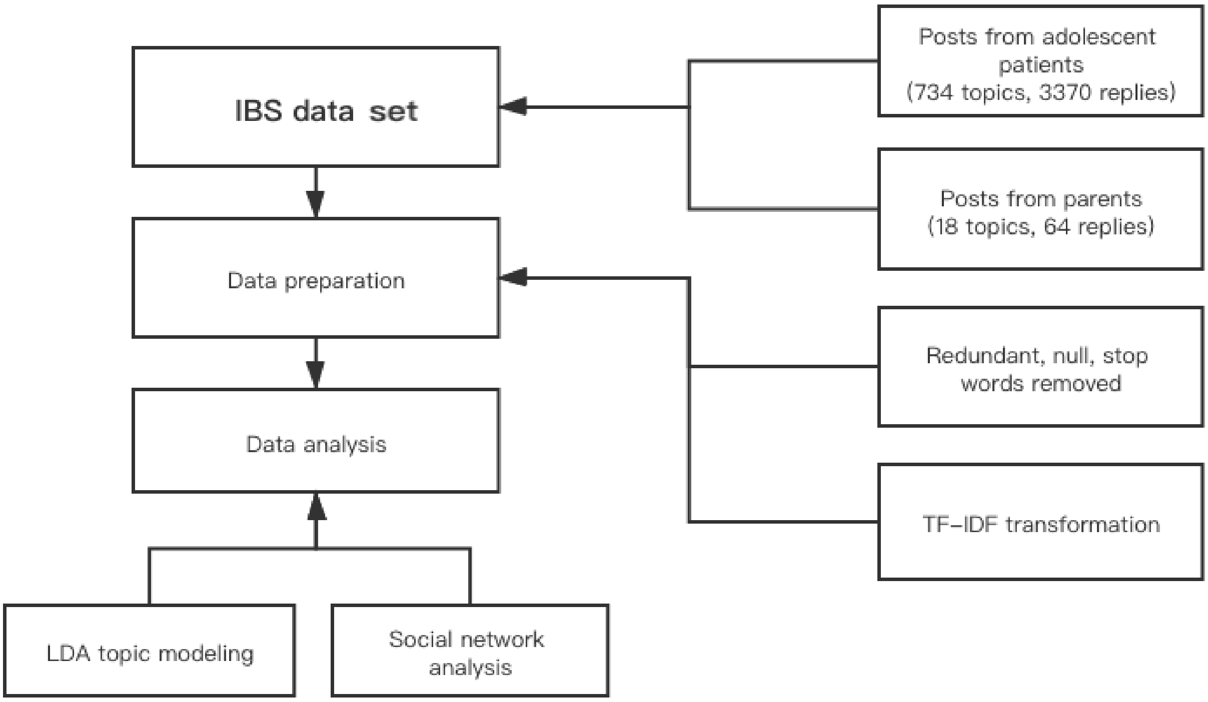
Sampling

In this study, we analyzed the topics in web posts from 2010 to 2019 from adolescents with IBS and the parents' posts concerning the IBS symptoms experienced by their teenaged daughters and sons. The data for this study were retrieved in June 2019 from the subforum discussing *Teens and Children's Issues* on the IBS Group forum [17]. All 734 topics in 3370 replies posted were collected and analyzed for adolescent patients. A preliminary analysis centers on identifying the topics when parents discussed the IBS issues experienced by their children aged between 13 and 17 years [18], involving 18 forum topics with 64 replies in this subforum. The LDA approach identified 6 major distinctive topics in their web-based discussion, as the analyses showed that if the number of topics was greater than 6, significant overlap could occur among them in the corpora.

Social Network Analysis

For social network analysis, we parsed all posts to identify the links that arose when a pair of users participated in a discussion about a subject on the same thread. These links were further dissected to identify the most active users on the forum and the nature of the interactions among them. We also plotted the social network graphs for this data set and performed further analyses. To clean the data, we first removed the null and duplicate posts. Then, we also deleted common stop words such as *this*, *that*, *a*, and *the*. We first built a document-term matrix (eg, an N×M document-term matrix D is a matrix with N documents and M terms). We then applied term frequency-inverse document frequency to process the data so as to eliminate the weight for high-frequency words appearing in every document and add weight for words appearing only in a few documents. Scikit-learn and pyLDAvis libraries were then used to perform topic modeling analyses in Python. Figure 1 shows the data collection and analysis process.

Figure 1. The process of data collection and analysis. IBS: irritable bowel syndrome; LDA: latent Dirichlet allocation; TF-IDF: term frequency-inverse document frequency.



Results

Overview

In this section, we first report the results from the *Parent's Discussion* subforum and then those from the *Teens' and Children's Issues* forum.

Parents' Perspectives on Their Children's IBS Issues

In the subforum of *Parent's Discussion* under the *Teens and Children's Issues* forum [18], we obtained 18 forum topics (or threads) and 64 replies after removing the duplicate, null, and invalid posts. A total of 82 posts were finally retrieved, including several comments from parents in the main forum: *Teens and Children's Issues*. Hence, we wanted to separate the parents' posts from those of the children. Parents' posts usually refer to their children as daughters, sons, he, and she. By looking for such posts, it is possible to determine whether they were made by a parent or caregiver and also the sex of their children. On the other hand, posts by children or teenagers usually do not refer to their own sex and, with some exceptions, it is usually not possible to determine the sex of the person making the post. By searching for sex-specific keywords, we identified and manually transferred 455 posts from the main forum into the parents' discussion subforum so that there were a total of 537 posts for analysis. The 6 topics that were discovered with 10 keywords each are listed in Table 1.

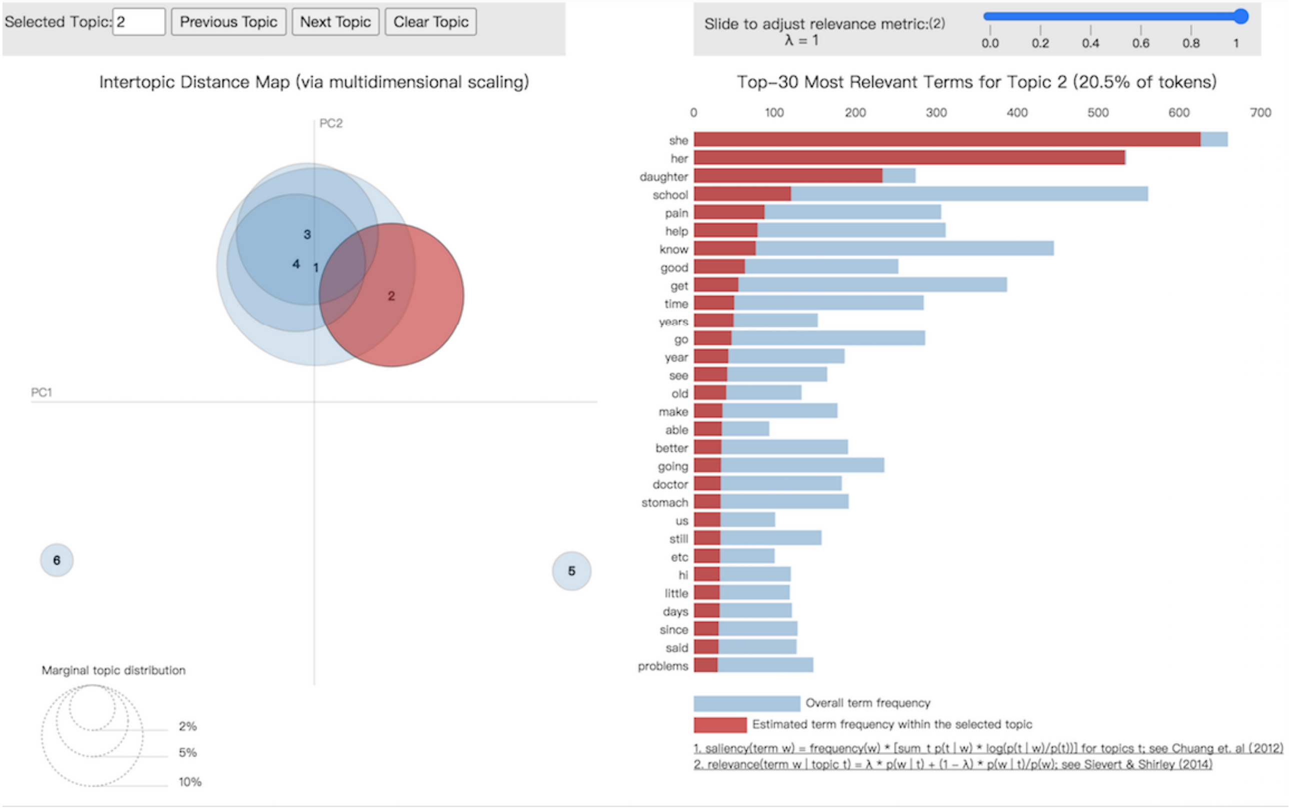
We used the keywords to assign appropriate topic names to each detected topic. We also used the modified pyLDAvis library to generate an overview and visualization of the topic distribution regarding girls' issues (Figure 2) and boys' issues (Figure 3). In the figures, the sizes of the circles represent the relative importance of topics and the distances between the topics reflecting their similarity to one another [19].

Table 1. The top 6 topics in the parents' discussion of their children's irritable bowel syndrome issues.

Topic number	Topic[a]	Keywords
1	School life	*school, know, like, feel, people, think, time, stomach, want, tell*
2	Girls' issues	*she, her, daughter, school, pain, help, know, good, get, time*
3	Boys' issues	*he, son, his, pain, school, help, year, doctor, old, know*
4	Diet choice	*foods, diet, like, eat, food, try, lot, gas, good, help*
5	Symptom	*person, she, say, everything, problem, better, life, week, never, symptoms*
6	Stress	*worse, think, keep, diet, stress, pain, bad, say, started*

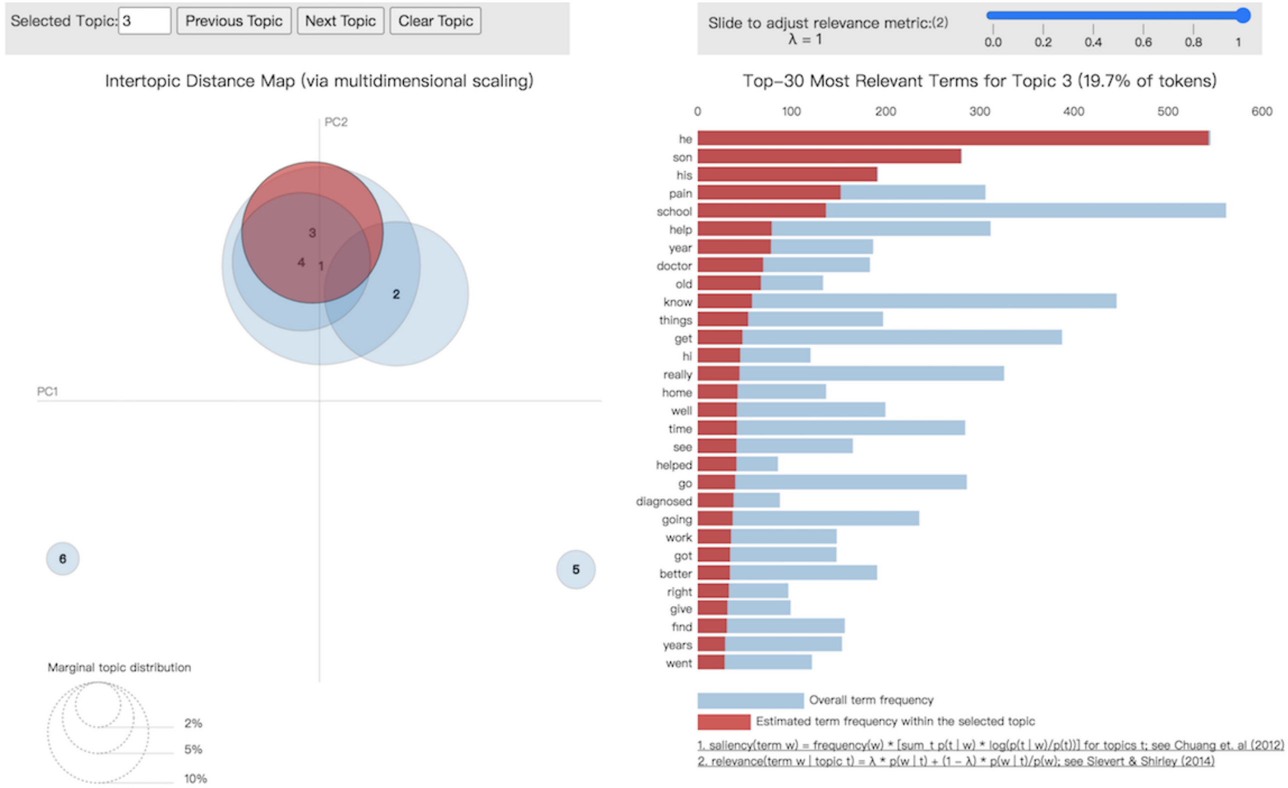
[a] The topic order is based on the total number of posts and replies each topic had from 1 (highest) to 6 (lowest).

Figure 2. Latent Dirichlet allocation visualization of topic 2 (girls' issues) in parents' subforum.



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 – λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Figure 3. Latent Dirichlet allocation visualization of topic 3 (boys' issues) in parents' subforum.



As shown in Figures 2 and 3, for parents, their adolescent girls' and boys' issues related to IBS fell into two major topic groups. The girls' issues, according to keywords representing the topic, are closely related to their school performance. Ignoring the frequently used keywords such as *she*, *her*, *daughter*, the other keywords indicate that parents seem to worry most about *school behavior* and the need of their children for *help*. Parent concerns about school attendance also appeared in their discussion on the subforum. An example post is as follows:

> *I know where we live, the schools get paid by the state based on how many days the kids are in school. So, this is why they are so concerned about the missed days. If I were you I would get a note from the doctor and bring it down to the school in person. Give a copy to the Principal, the Teacher and the School Nurse and explain the situation to them. This should help both your daughter and the school. Hopefully they will become a little more sympathetic to your daughter's needs and she will feel better about going to school overall. Don't let them bully you. Best of Luck and I hope your daughter starts feeling better soon.*

Furthermore, parental worries about their daughter missing classes appeared in the discussion, as indicated by the following post:

> *See if you can get a 504. I have a similar problem, and when I first asked to start a 504, the school counselor said she is fine and doing well and doesn't need one. She has not gotten into trouble for missing school, however, it causes both of us a lot of stress. I will be working on getting her a 504 next week so she will have it in place for high school. (A 504 is a plan for students who don't qualify for an IEPER but still have "disabilities" that impair their normal every day activities.)*

In contrast to the parents' worries about their daughters' school performance, for their sons, they appeared to be most concerned about their experience of pain and other symptoms brought on by IBS. This type of reaction can be seen in the discussion through posts such as the following:

> *My boy had a very strong reaction to the fructose solution. But none to the lactose. This was the only test in the long litany of tests around this miserable thing we call IBS that actually came out with an abnormal result. As a result he is on a low fructose and no fructan (fructans are in wheat and in the onion family especially). He has Post infectious IBS which was very severe during the first year of having it (he was treated by a chronic pain team in the end). Currently he has months when he is OK, normal really then the bowel problems always come back after a virus involving a sore throat.*

Some other examples of parents discussing treatment of their child's pain problems are as follows:

> *He is now on a very limited diet but his symptoms are so much better. He no longer has diarrhea or constipation on a regular basis, his stomach is no longer distended and best of all he hardly ever complains of stomach pains. We were also able to take him off all medications which treated symptoms but were not treating the cause of the pain. I suggest you see a good GI and have her tested for lactose*

> *intolerance and fructose malabsorption through Breath test and consider and endoscopy. Hope this helps and best of luck. I know it can be horrible to see your child in pain and feel you can't do much about it.*

> *My son has had for over 2 weeks pain in his stomach and usually in the middle and he goes to the bathroom but not a lot. It's not hard and not really diarrhea but I think he has been backed up. He has had nausea and a headaches with this. He either is really hungry or not into eating at all.*

Some other keywords also appeared in the discussion related to boys, such as *doctor* indicating parents' excessive attention to doctors' advice, diagnosis, and treatment, as it relates to boys. For example:

> *Now if your son is going to attend school great for him! I have a couple tips that might help him out with bathrooms etc. Go first to your doctor or specialist, have them fill out a form for IBS or a Doctor's note that say the YES he does have IBS.*

Adolescent Patients' Perspectives on Their IBS Issues

In the *Teens and Children's Issues* forum [17], discussion occurred among the adolescents who were themselves the patients. We crawled 734 topics (or threads) with 3370 replies on this forum. After deleting the null and duplicate data, 2410 comments remained. Then, we removed the former 455 parents' comments, which were transferred to the parents' discussion; thus, the comments of the 1955 teenagers and children were left for further analysis. As done before, we cleaned the data and applied the LDA method. This generated 6 main topics, which are listed in Table 2. We also created visualizations for the topics as shown in Figure 4.

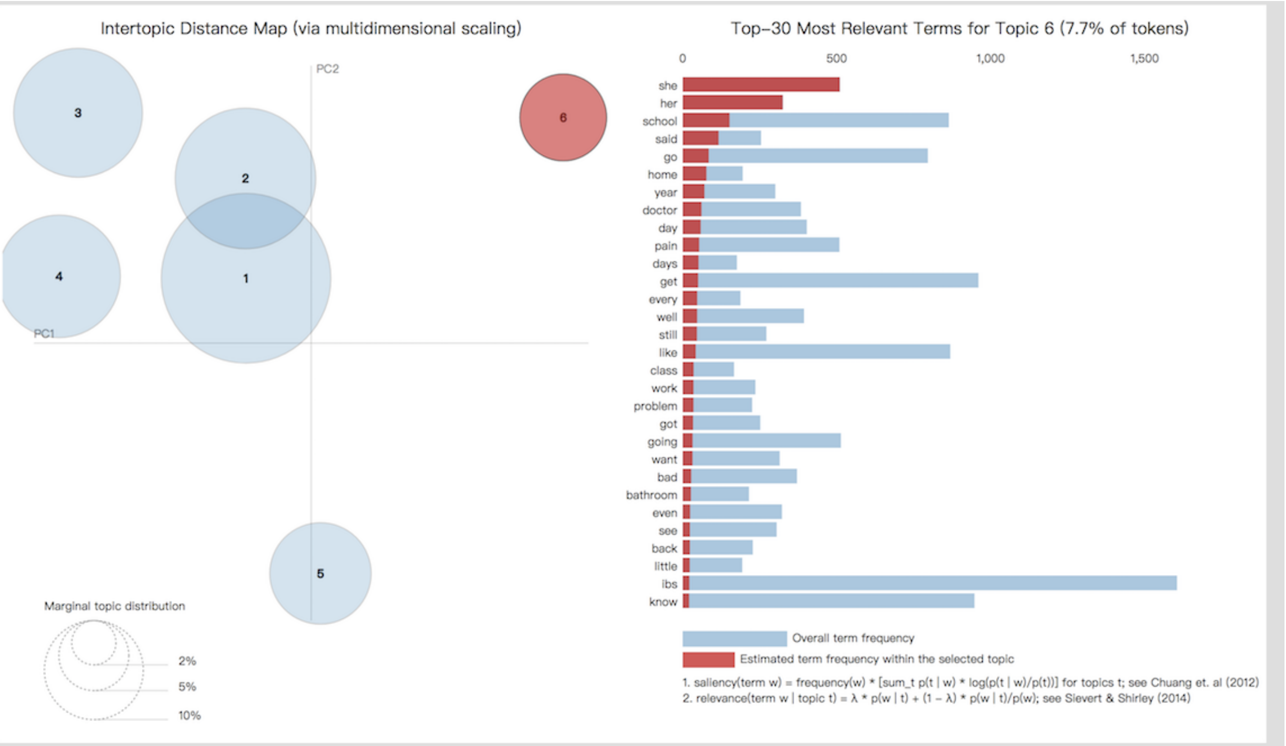
Table 2. The top 6 topics in the adolescent patients' discussion of their irritable bowel syndrome issues.

Topic number	Topic[a]	Keywords
1	School life	*IBS*[b], *school*, *know*, *life*, *try*, *think*, *hard*, *stress*, *want*, *need*
2	Treatment or diet	*eat*, *day*, *time*, *stomach*, *bad*, *IBS*, *sick*, *think*, *lot*, *little*
3	Symptoms	*IBS*, *pain*, *symptoms*, *diet*, *try*, *foods*, *help*, *food*, *find*, *thanks*
4	Boys' ties to doctors	*he*, *help*, *doctor*, *feel*, *need*, *pain*, *know*, *IBS*, *try*
5	Social or friend issues	*friends*, *know*, *school*, *feel*, *talk*, *understand*, *time*, *hope*, *think*
6	Girls' ties to doctors	*she*, *her*, *school*, *home*, *doctor*, *pain*, *class*, *work*, *problem*

[a] The topic order is based on the total number of posts and replies each topic had, ranging from 1 (highest) to 6 (lowest).

[b] IBS: irritable bowel syndrome.

Figure 4. Latent Dirichlet allocation visualization of the top 6 topics in the posts from adolescent patients.



As shown in Table 2, we concluded based on the LDA topic modeling analysis that there were 6 main topics in the posts. The first topic was calculated with keywords such as *IBS*, *school*, *know*, *life*, *try*, *think*, *hard*, and *stress*. We named this topic *school life*. The second topic was calculated with keywords such as *eat*, *day*, *time*, *stomach*, *bad*, *IBS*, *sick*, and *think*, and we named this topic *treatment or diet*. The third topic was calculated with keywords such as *IBS*, *pain*, *symptoms*, *diet*, and *try foods*, and we named this topic *symptoms*. It should be noted that schools, symptoms, and treatments were the most mentioned topics.

Among the symptoms, keyword frequencies showed that *stomach* and *pain* issues are the most devastating and irritating symptoms bothering patients in their daily lives. It is evident that for proper treatment, *food* and *diet* must be carefully planned. Eating more or less, quantity and diet choices are most frequently mentioned as ways to alleviate the problem in the diet options–related discussion. However, different patients have very personalized situations that they must pay attention to. Some discussed whether chocolate is suitable for patients with IBS, for example:

> *I feel your pain on not eating chocolate, but I'm so used to not getting to eat good food' anymore that I just kind of pass right by it and don't notice. Others concluded, on no junk food, for example: "no junk food, fried foods, dairy products, chocolate, soda...nothing!" Some patients seek help from doctors, some get advice from nutritionist, suggesting that "Often times it's your diet, and a nutritionist can really help you in figuring out what foods are causing these problems."*

School performance was also an important topic for young patients and a cause of stress. They encountered difficulty in *fitting* into society, some mentioning experiences such as:

> *In high school I missed 1.5-2 days a week, and it was even more embarrassing because people noticed I frequently took days off and when I did show up the teachers, students used to brazenly point out my attendance (good to see you! followed by laughter). I avoided school even more. I wanted so desperately to graduate and not be forced to sit in a classroom.*

Evidently, it was hard for them to be understood by their peers and teachers in school, with some complaining as follows:

> *Missing school is the worst as I know that nobody really understands what I'm going through.*

One theme consistent in both the parents' and teens' forums is that although both worried and suffered stress or anxiety from IBS, their emphases differed. For parents, the stress was mostly related to IBS symptoms, such as pain, whereas for teenagers and children, it came from school and social activities. This means that parents should be encouraged to become more closely involved in the psychological and social impact of IBS on their children's daily lives.

Social Network Analysis

The purpose of the social network analysis was to understand the patterns of interaction among the users on this forum and extract insights from them. An interaction occurs between a pair of users when both are involved in the same thread by making posts related to the same topic, say, in response to a question raised on the forum. We found that there were 3010 unique interactions in all the posts on the *Teens and Children's Issues* forum, including the parents' subforums. After excluding the responses by the same author who created the thread, there were 2053 interactions. On examining the data more closely, we found that user *Kathleen M*, one of the top content contributors, was an IBS expert who also participated in the IBS forum for adults.

Next, we plotted the top 30 most active users who contributed the most to the forum (Figure 5). The plot shows a typical power law pattern, where most of the contributions come from a small number of users and the participation of the remaining users is very small. This is also consistent with the pattern observed for the adult forum in our previous study [20,21].

Next, we used the Gephi tool [22] to create a social network graph of all interactions (Figure 6). This graph shows that there is a strongly connected core at the center from which there are connections radiating to the other nodes. The nodes on the outside have weak connections with each other. In this graph, there are 1165 nodes and 2020 edges. The edges are weighted by the number of exchanges that occur between the pair of users on the 2 sides. The weights vary from 1 to 5. The graph is undirected, so an edge with weight 1 represents the flow of a message in both directions. The density of the graph is 0.001, and its diameter is 16. The average degree of this graph is 1.734, and the weighted degree is 2.411. The clustering coefficient is 0.034.

Figure 5. Plot of number of posts made by the 30 top content contributors.

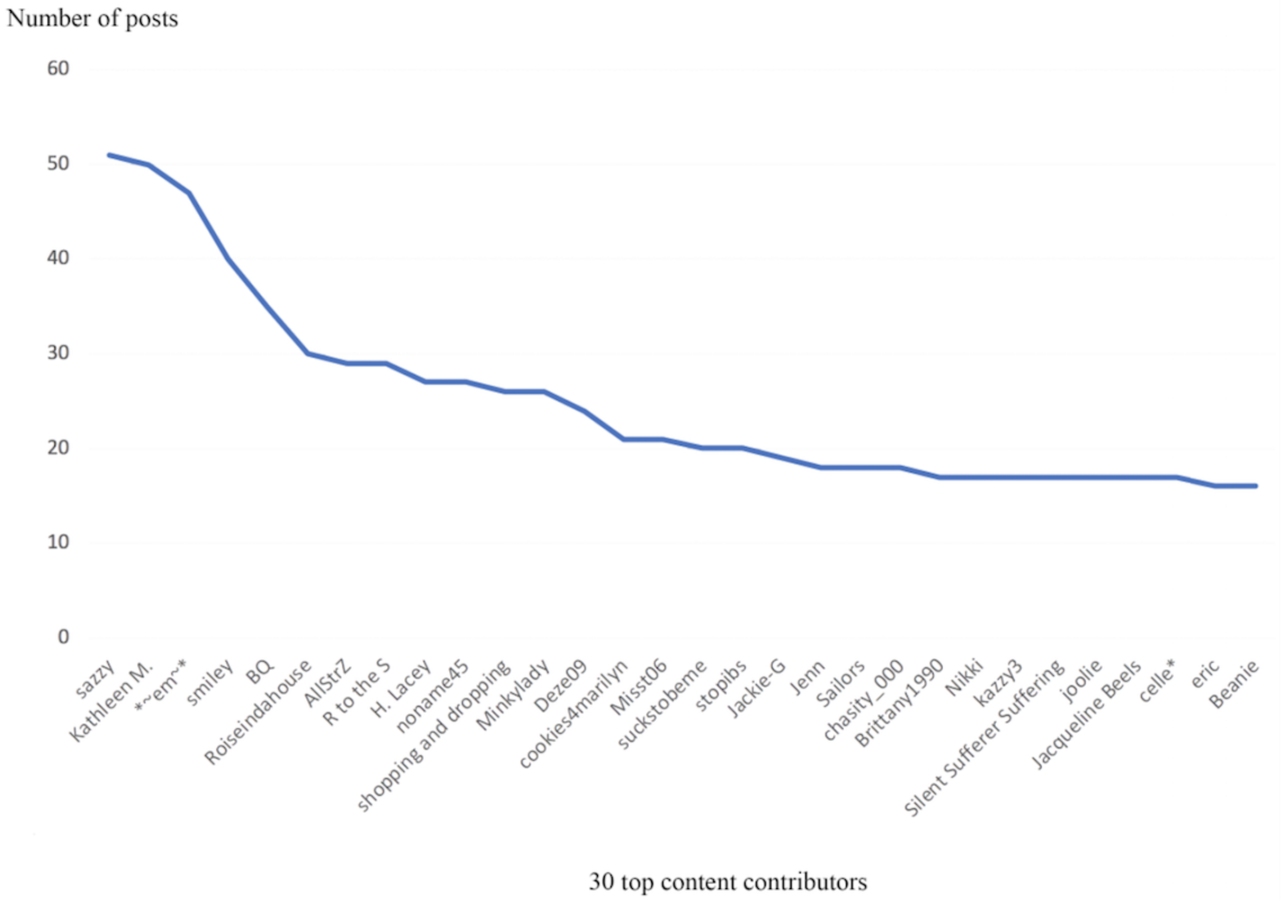
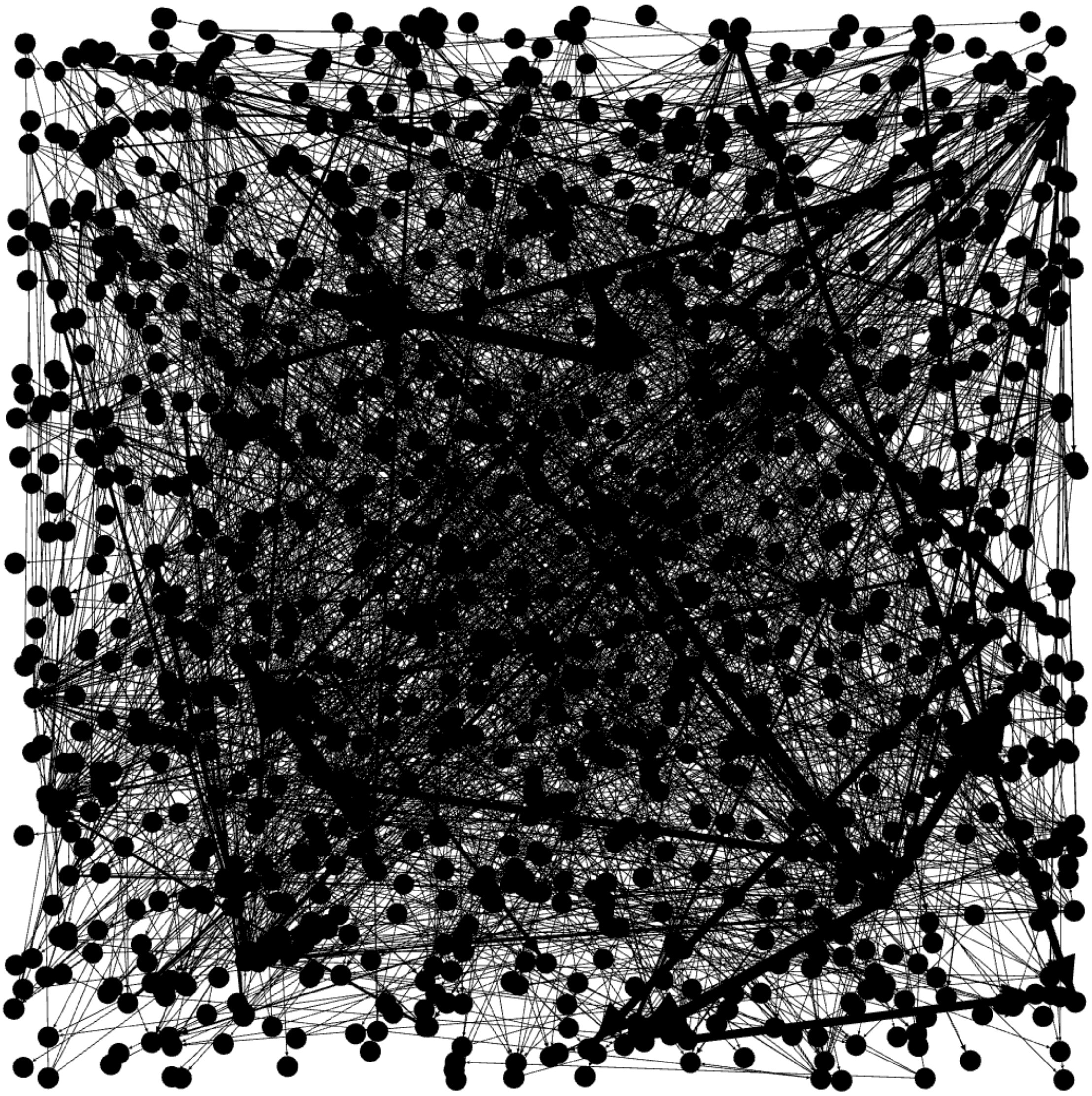


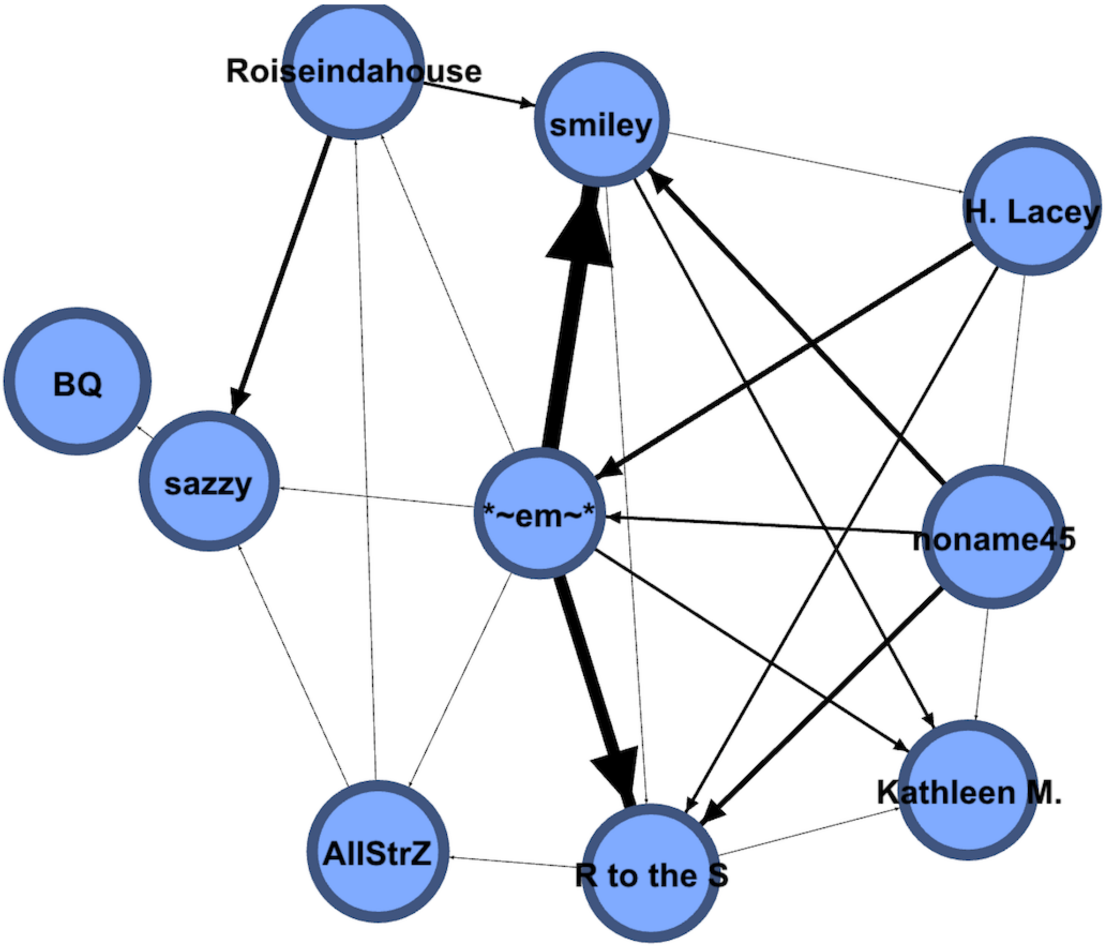
Figure 6. A full social network for all 2053 interactions on the forum.



We also show a smaller filtered version of this graph in Figure 7. We deleted self-connected edges and showed a graph with nodes of degree >25 using the filter feature of Gephi to focus on the links among the most prolific contributors. The degree of a node in a social network is the number of connections it has to other nodes, and the degree distribution is the probability distribution of these degrees over the entire network, including both out degree, or the number of edges leaving a vertex, and in degree, which is the number of edges entering a vertex. The analyses show that this is a more tightly connected graph, as expected. The average degree of this graph and the average weighted degree are 2.6 and 4.867, respectively. The diameter is much smaller at 5, and the density is higher at 0.186. The clustering coefficient is 0.233.

As Figure 7 shows, the top contributors to the content in the forum had more interactions among themselves than with all other users in the same network. Among the top contributors, *Kathleen M* was found to contribute the most. It is important to note that this user or node did not have the highest closeness centrality, which reflects the inverse of one node’s average distance to all other nodes [23] and is one of the most important centrality indices in a network for the nodes [24]. Rather, *Kathleen M* was the top commentator on posts of others who did not post any inquiries in the forum. Other top content contributors, such as **~em~** and **sazzy**, posted inquires and comments on others’ messages and had the highest betweenness centrality in the forum, which is a key measure of distance in a network for a node that is calculated as the number of shortest paths going through a node. For **~em~**, the degree is 62, the closeness centrality is 0.275, and the betweenness centrality is 11,816.665, whereas for **sazzy**, the degree is 54, the closeness centrality is 0.168, and the betweenness centrality is 9749.487. These users were the 2 nodes with the highest degrees, closeness centrality scores, and betweenness centrality scores, even higher than those of the top content producer *Kathleen M* (degree=50, closeness centrality=0, and betweenness centrality=0). The results show that *Kathleen M* replies to many posts but did not post in this forum section, whose out degree is 0 and in degree is 50.

Figure 7. The social network of users with high-degree nodes (degree >25).



Discussion

Principal Findings

This study has discovered that IBS can be a sensitive topic for younger patients, particularly adolescents who are already undergoing myriad other critical developmental changes that deter them from sharing their feelings and emotions with adults, including their parents, caregivers, and teachers. Given the increasingly dominant role played by web-based platforms as a source of information exchange in this digital age, especially among adolescents [25], it is critical to harness information from web-based patient- and caregiver-centered forums to better understand key challenges, concerns, and issues of interest to younger patients with IBS and their caregivers; examine how health information, including misinformation and obsolete health care practices, is propagated and perpetuated via web media; and how the rates of information propagation and dissemination depend on the sentiments underlying health-related messages and links among forum members.

To date there is no definite cure for IBS, and hence, health care forums for IBS are very helpful to adult patients in exchanging medical information about their conditions and obtaining emotional support from others. Our semantic analysis of the posts reveals that adolescent patients with IBS and their parents also benefit from raising their issues and concerns regarding a forum dedicated to them. Physicians often advise them to follow a diet low in fermentable oligosaccharides, disaccharides, monosaccharides, and polyols and avoid *trigger foods* [13]. It is also important for teenagers to learn how to manage pain and stress, which are known to worsen the symptoms [26].

Previous research has found that adult patients tend to focus on gastrointestinal discomfort in their posts, and the top three causes were psychological factors, food, and allergens [27]. Using the LDA algorithm and an LDA visualization tool, we identified and visualized 6 main topics in our study from 2 health forums, one for discussion among the parents who cared for teen patients with IBS and another for discussion among the adolescent patients themselves. For parents, a top concern was the school performance of their daughters and how much *help* they got at school. For their sons, they were more concerned about the pain and suffering they had to endure. Teenagers themselves are most concerned about the effect of IBS on their everyday activities and social lives.

Although both sexes faced worries and anxieties from IBS, boys and girls responded differently to IBS symptoms, especially abdominal pain. It is evident that both teenagers with IBS and their parents valued the social support they received from web-based health care forums enormously. Our findings should also assist medical professionals and school administrators in enacting better health policies and lead to sound decisions and higher care quality for patients with IBS. A better understanding of the impact of IBS on young patients' social and school lives can provide powerful support to both adolescents' caregivers in self-caring IBS symptoms. Such self-management techniques based on the information shared on health forums are known as *a web intervention*, which have positive outcomes for patients with IBS [28]. Our research adds fresh evidence of the effectiveness of such a web intervention.

Going beyond previous studies on patients with IBS that seldom analyzed sex differences, we identified three reasons for the sex differences that were observed on the forum. First, girls found it difficult to describe their IBS issues to others to seek help and support. Some even became more reticent and felt humiliated easily. For instance, one adolescent was asked in front of her parents whether she was pregnant. Second, some young women have been brought up by parents to follow social norms and expectations, more so than boys. Third, the tolerance of pain varies between females and males with girls being more capable of enduring pain than boys. The findings reveal that in addition to coping with IBS symptoms, young patients have to face the challenges of social influences and anxiety associated with this health disorder in addition to physical pain and other symptoms. Boys and girls are affected differently by pain and school performance, and their views are different from those of parents.

Our social network analyses have identified multiple prolific users in the forum who contributed most content, such as posts and replies, in the forum. They functioned like opinion leaders in the network that could strongly influence other users' health information processing concerning IBS. Their posts and replies could be the major drivers that kept the conversation of adolescent IBS symptoms going among the users, which also added value to the information flow. It is important to note that this study has found that one of the top content contributors such as *Kathleen M* did not post any messages but commented and replied to others' posts. Without the social network analysis, it was impossible to know that the top content contributor achieved the status by writing replies and comments only, who was more influential than many other users who posted inquires and comments about IBS. Other influential users in this network, such as users *sazzy*, **~em~**, and *smiley*, posted inquires and commented on messages from their peers. This finding has important implications for users and caregivers when they exposed, accessed, and accepted the health information circulated in the forum. Without this study, average users would not know the veiled features of the information flow from the top contributors in the health forum.

Several limitations must be considered when interpreting these findings. First, the results may not be representative of the entire population of adolescents with IBS even when this study analyzed all the posts and replies from adolescent patients and parents on the IBS Group forum. Second, we cannot recognize the sex of the parents based on their posts and can only indirectly identify some of the adolescents' sex by pronouns used, which may not be completely accurate. Third, because this is an anonymous forum, little information can be collected about the users, such as age, school year, health history including diagnoses, medicines, tests, allergies, immunizations, and treatment plans. All these factors make it difficult to fully understand their medical conditions, and thus, only limited insights may be revealed.

Conclusions

IBS among children and teenagers is an underappreciated problem in the medical and school communities, although about 1 of every 7 adolescents suffers from it. This research uses a novel machine learning approach called the LDA algorithm to perform topic analysis on 2 forums—one for parents and another for teens to discuss their IBS problems. One of the important findings is that certain IBS concerns in teenagers are sex specific, which is not so apparent among adult patients. Parents have different issues that bother them for their sons or daughters. Some issues among girls and boys who suffer from this condition are also different. Both parents and children derive considerable medical and emotional benefits from IBS forums. Our study shows that this condition is very debilitating for adolescents, and in many cases, it has a devastating effect on their everyday activities and social lives with potentially lifelong consequences. This study represents the first attempt to leverage both machine learning approaches and social network analysis to identify top IBS concerns from the perspectives of adolescent patients and caregivers on the same web-based health forums. The insights revealed in this study, especially the social influences of IBS-related concerns, should help adolescent patients manage their symptoms and school lives. Moreover, caregivers, teachers, and school administrators should also benefit from the findings of this study.

Conflicts of Interest

None declared.

References

1. Hoyle B. IBS in Children and Teens. Everyday Health. URL: https://www.everydayhealth.com/ibs/ibs-children-and-teens.aspx [accessed 2021-05-20]
2. Hyams JS, Burke G, Davis PM, Rzepski B, Andrulonis PA. Abdominal pain and irritable bowel syndrome in adolescents: a community-based study. J Pediatr 1996 Aug;129(2):220-226. [doi: 10.1016/s0022-3476(96)70246-9] [Medline: 8765619]
3. Yang F, Zhong B, Kumar A, Chow S, Ouyang A. Exchanging social support online: a longitudinal social network analysis of irritable bowel syndrome patients' interactions on a health forum. J Mass Commun Q 2017 Sep 28;95(4):1033-1057. [doi: 10.1177/1077699017729815]
4. Zhao J, Han H, Zhong B, Xie W, Chen Y, Zhi M. Health information on social media helps mitigate Crohn's disease symptoms and improves patients' clinical course. Comput Hum Behav 2021 Feb;115:106588. [doi: 10.1016/j.chb.2020.106588]

5. Ouyang A, Inverso NA, Chow S, Kumar A, Zhong B. Mo1639 'listening' to IBS patients in the 21st century: offerings from an on-line self help and support group. Gastroenterology 2016 Apr;150(4):S739. [doi: 10.1016/s0016-5085(16)32510-0]
6. Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinformatics 2015;16(Suppl 13):S8 [FREE Full text] [doi: 10.1186/1471-2105-16-S13-S8] [Medline: 26424364]
7. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022.
8. Saps M, Seshadri R, Sztainberg M, Schaffer G, Marshall BM, Di Lorenzo C. A prospective school-based study of abdominal pain and other common somatic complaints in children. J Pediatr 2009 Mar;154(3):322-326. [doi: 10.1016/j.jpeds.2008.09.047] [Medline: 19038403]
9. Caplan A, Walker L, Rasquin A. Validation of the pediatric Rome II criteria for functional gastrointestinal disorders using the questionnaire on pediatric gastrointestinal symptoms. J Pediatr Gastroenterol Nutr 2005 Sep;41(3):305-316. [doi: 10.1097/01.mpg.0000172749.71726.13] [Medline: 16131985]
10. Miele E, Simeone D, Marino A, Greco L, Auricchio R, Novek SJ, et al. Functional gastrointestinal disorders in children: an Italian prospective survey. Pediatrics 2004 Jul 1;114(1):73-78. [doi: 10.1542/peds.114.1.73] [Medline: 15231910]
11. Coulson NS. Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. Cyberpsychol Behav 2005 Dec;8(6):580-584. [doi: 10.1089/cpb.2005.8.580] [Medline: 16332169]
12. Self MM, Czyzewski DI, Chumpitazi BP, Weidler EM, Shulman RJ. Subtypes of irritable bowel syndrome in children and adolescents. Clin Gastroenterol Hepatol 2014 Sep;12(9):1468-1473 [FREE Full text] [doi: 10.1016/j.cgh.2014.01.031] [Medline: 24486406]
13. Devanarayana NM, Rajindrajith S. Irritable bowel syndrome in children: current knowledge, challenges and opportunities. World J Gastroenterol 2018 Jun 7;24(21):2211-2235 [FREE Full text] [doi: 10.3748/wjg.v24.i21.2211] [Medline: 29881232]
14. Chey WD, Kurlander J, Eswaran S. Irritable bowel syndrome: a clinical review. J Am Med Assoc 2015 Mar 3;313(9):949-958. [doi: 10.1001/jama.2015.0954] [Medline: 25734736]
15. Carlson MJ, Moore CE, Tsai CM, Shulman RJ, Chumpitazi BP. Child and parent perceived food-induced gastrointestinal symptoms and quality of life in children with functional gastrointestinal disorders. J Acad Nutr Diet 2014 Mar;114(3):403-413 [FREE Full text] [doi: 10.1016/j.jand.2013.10.013] [Medline: 24360501]
16. Silipo R. LDA for text summarization and topic detection. DZone-AI Zone 2019:-.
17. Teens and Children Issues. IBS Group. URL: https://www.IBSgroup.org/forums/forum/26-teens-and-childrens-issues [accessed 2021-05-20]
18. Prent Discussion. IBS Group. URL: https://www.IBSgroup.org/forums/forum/95-parents-discussion [accessed 2021-05-20]
19. Chuang J, Ramage D, Manning C, Heer J. Interpretation and Trust: Designing Model-driven Visualizations for Text Analysis. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI'12; May 12-14, 2012; Austin, Texas. [doi: 10.1145/2207676.2207738]
20. Kumar A, Zhao K. Making Sense of a Healthcare Forum - Smart Keyword and User Navigation Graphs. In: Thirty Fourth International Conference on Information Systems. 2013 Presented at: CIS'13; December 2-5, 2013; Milan, Itlay.
21. Chow S, Kumar A, Ouyang A, Zhong B, Lee J, Inverso N. What Can Physicians Learn From Social Forums: Insights From an Online Self Help and Support Group. In: Proceedings of the IEEE 7th International Conference on Computational Advances in Bio and Medical Sciences. 2017 Presented at: ICCABS'17; October 19-21, 2017; Orlando, FL p. 19-21. [doi: 10.1109/iccabs.2017.8114302]
22. Gephi - The Open Graph Viz Platform. 2020. URL: https://gephi.org/ [accessed 2021-06-01]
23. Sabidussi G. The centrality of a graph. Psychometrika 1966 Dec;31(4):581-603. [doi: 10.1007/BF02289527] [Medline: 5232444]
24. Bavelas A. A mathematical model of participation in small group discussions. Behav Sci 1965 Apr;10:164-166. [doi: 10.1002/bs.3830100206] [Medline: 14284291]
25. Sandhu BK, Paul SP. Irritable bowel syndrome in children: pathogenesis, diagnosis and evidence-based treatment. World J Gastroenterol 2014 May 28;20(20):6013-6023 [FREE Full text] [doi: 10.3748/wjg.v20.i20.6013] [Medline: 24876724]
26. Zemel L, Blier PR. Juvenile Fibromyalgia: A Primary Pain, or Pain Processing, Disorder. Semin Pediatr Neurol 2016 Aug;23(3):231-241. [doi: 10.1016/j.spen.2016.10.007] [Medline: 27989331]
27. Schäfer F, Faviez C, Voillot P, Foulquié P, Najm M, Jeanne J, et al. Mapping and Modeling of Discussions Related to Gastrointestinal Discomfort in French-Speaking Online Forums: Results of a 15-Year Retrospective Infodemiology Study. J Med Internet Res 2020 Nov 03;22(11):e17247 [FREE Full text] [doi: 10.2196/17247] [Medline: 33141087]
28. Tonkin-Crine S, Bishop FL, Ellis M, Moss-Morris R, Everitt H. Exploring patients' views of a cognitive behavioral therapy-based website for the self-management of irritable bowel syndrome symptoms. J Med Internet Res 2013 Sep 03;15(9):e190 [FREE Full text] [doi: 10.2196/jmir.2672] [Medline: 24001787]

Abbreviations

IBS: irritable bowel syndrome
LDA: latent Dirichlet allocation

RQ: research question

Edited by R Kukafka; submitted 31.12.20; peer-reviewed by Y Zhu, T Taft; comments to author 16.03.21; revised version received 14.04.21; accepted 06.05.21; published 09.06.21.

Please cite as:
Zhong B, Liu Q
Medical Insights from Posts About Irritable Bowel Syndrome by Adolescent Patients and Their Parents: Topic Modeling and Social Network Analysis
J Med Internet Res 2021;23(6):e26867
URL: https://www.jmir.org/2021/6/e26867
doi:10.2196/26867
PMID:34106078

Original Paper

A Web Interface for Antibiotic Prescription Recommendations in Primary Care: User-Centered Design Approach

Ronni Madar[1], MD; Adrien Ugon[2,3], PhD; Damir Ivanković[4], MD; Rosy Tsopra[1,5,6,7], MD, PhD

[1]Université Sorbonne Paris Nord, Bobigny, France

[2]ESIEE-Paris, Noisy-le-Grand, France

[3]Laboratoire d'Informatique de Paris 6, CNRS, Sorbonne Université, Paris, France

[4]Department of Public and Occupational Health, Amsterdam Public Health Research Institute, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands

[5]Inserm, Université de Paris, Sorbonne Université, Centre de Recherche des Cordeliers, Information Sciences to support Personalized Medicine, Paris, France

[6]Inria Paris, Paris, France

[7]Department of Medical Informatics, Hôpital Européen Georges-Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

Corresponding Author:
Rosy Tsopra, MD, PhD
Inserm, Université de Paris, Sorbonne Université, Centre de Recherche des Cordeliers, Information Sciences to support Personalized Medicine
15 rue de l'école de médecine
Paris, 75006
France
Phone: 33 56092167
Email: rosy.tsopra@nhs.net

Abstract

Background: Antibiotic misuse is a serious public health problem worldwide. National health authorities release clinical practice guidelines (CPGs) to guide general practitioners (GPs) in their choice of antibiotics. However, despite the large-scale dissemination of CPGs, GPs continue to prescribe antibiotics that are not recommended as first-line treatments. This nonadherence to recommendations may be due to GPs misunderstanding the CPGs. A web interface displaying antibiotic prescription recommendations and their justifications could help to improve the comprehensibility and readability of CPGs, thereby increasing the adoption of recommendations regarding antibiotic treatment.

Objective: This study aims to design and evaluate a web interface for antibiotic prescription displaying both the recommended antibiotics and their justifications in the form of antibiotic properties.

Methods: A web interface was designed according to the same principles as e-commerce interfaces and was assessed by 117 GPs. These GPs were asked to answer 17 questions relating to the usefulness, user-friendliness, and comprehensibility and readability of the interface, and their satisfaction with it. Responses were recorded on a 4-point Likert scale (ranging from “absolutely disagree” to “absolutely agree”). At the end of the evaluation, the GPs were allowed to provide optional, additional free comments.

Results: The antibiotic prescription web interface consists of three main sections: a clinical summary section, a filter section, and a recommended antibiotics section. The majority of GPs appreciated the clinical summary (90/117, 76.9%) and filter (98/117, 83.8%) sections, whereas 48.7% (57/117) of them reported difficulty reading some of the icons in the recommended antibiotics section. Overall, 82.9% (97/117) of GPs found the display of drug properties useful, and 65.8% (77/117) reported that the web interface improved their understanding of CPG recommendations.

Conclusions: The web interface displaying antibiotic recommendations and their properties can help doctors understand the rationale underlying CPG recommendations regarding antibiotic treatment, but further improvements are required before its implementation into a clinical decision support system.

(J Med Internet Res 2021;23(6):e25741) doi:10.2196/25741

KEYWORDS

clinical decision support system; visualization; usability; clinical practice guidelines; antibiotic; primary care

Introduction

Antibiotic misuse is a serious public health problem worldwide [1,2]. It exposes patients to the risk of adverse effects and complications, including death [3,4], as well as bacterial resistance [5]. Most antibiotic prescriptions are made in primary care settings. In this context, the choice of antibiotic is usually empiric (ie, without identification of the causative bacterium) and depends on various microbiological, epidemiological, pharmacological, patient condition–, and general practitioner (GP)-related factors [6].

National health authorities release clinical practice guidelines (CPGs) to guide GPs in their choice of antibiotics. CPGs contain evidence-based recommendations from a group of experts based on scientific publications. However, despite the large-scale diffusion of CPGs, GPs continue to prescribe antibiotics that are not recommended for first-line treatment (eg, broad-spectrum antibiotics) [7,8]. This noncompliance with recommendations may be due to the GPs misunderstanding the CPGs or their lack of confidence in these guidelines [9,10]. Indeed, it has been shown that GPs are suspicious of the content of CPGs; they believe that the recommendations are driven by economic issues rather than a desire to improve patient care and that there is a lack of evidence to support the recommendations [11]. They also find the guidelines unclear, ambiguous, incomplete, complex, and unusable in clinical practice [11].

An antibiotic prescription web interface displaying not only recommendations of antibiotics but also their justifications could help improve the GPs’ comprehensibility and readability of CPGs, thereby increasing the adoption of recommendations regarding antibiotic treatment. The justifications for antibiotic recommendations can be found within CPG documents, but they are often *lost* or *hidden* in large amounts of text [12,13]. A qualitative analysis of CPGs [12,13] for antibiotic treatment showed that these justifications were based on antibiotic properties [14,15]. For example, fosfomycin trometamol is recommended for the treatment of uncomplicated cystitis because of the following properties: short-duration protocol, few side effects, and little collateral damage. The display of antibiotic properties in an easily accessible and understandable manner could therefore help improve the comprehensibility and readability of CPGs for GPs, thereby increasing the chances of them successfully adopting those recommendations.

It is not easy to display the recommended antibiotics and their properties in a readily usable interface [16,17]. In the context of antibiotic treatment, several kinds of interface have been used, including textual formats [18,19], tables [19], diagrams [20], hypertextual links [21], and tick boxes [18]. Outside the domain of medicine, e-commerce interfaces (such as those used by e-commerce organizations like Booking, eBay, and Amazon) are widespread and make it possible for consumers to compare particular parameters between products. However, surprisingly, interfaces of this type are rarely used in the medical domain. Such e-commerce interfaces could be an effective way of displaying both antibiotics and their properties.

In this study, we aim to (1) design a web interface for antibiotic prescription presenting recommendations and justifications, in the form of an e-commerce interface displaying antibiotics and their properties, and (2) evaluate the readability and utility of this interface for improving the GPs’ comprehension of CPGs.

Methods

Interface Design

The elements involved in medical decision-making for antibiotic prescription have been previously identified based on analyses of clinical guidelines [14] and clinical expertise [6].

The best way to organize these elements into the proposed web interface was determined by reviewing and analyzing the content of the following:

1. Health care interfaces: However, no interface with a design similar to that of an e-commerce interface was retrieved.
2. Several well-known e-commerce websites for booking trips or hotels and web-based stores: A panel of websites were reviewed and the interface of four selected e-commerce websites (ie, Booking, Amazon, Tripadvisor, and LeBonCoin) were analyzed in greater detail. This analysis revealed that these interfaces contained three common main sections: a summary section, a result section, and a filter section (Figure 1). The same partitioned sections were then used to organize the elements involved in medical decision-making and to design the web interface.

The *summary* section of e-commerce interfaces usually displays the search criteria entered by the user when searching for a product (eg, “hotels in Paris from 21/12/2020 to 26/12/2020”). In our web interface, the elements “patient profiles” and “diseases” (eg, adult pharyngitis) were considered as search criteria.

The *filter* section of e-commerce interfaces usually displays optional parameters that users can click to filter the products selected in response to the initial search criteria (eg, hotel price). In our web interface, elements relating to certain generic patient conditions that are taken into account during drug prescription (eg, renal failure) were considered as optional parameters.

Finally, the *result* section of e-commerce interfaces usually displays the products recommended after both the search criteria and the optional parameters entered by the user have been taken into account. The characteristics of the products are usually also displayed in the form of icons to facilitate product comparison (eg, an icon for a swimming pool). In the web interface, the elements “antibiotic” and “antibiotic properties” were considered as products and characteristics, respectively. The antibiotics recommended in CPGs and the properties used by CPG experts to make those recommendations would then be displayed. A previous study identified seven preference properties (eg, “convenient protocol”) that are currently used by experts for making recommendations [14]. Antibiotics satisfying these properties are preferred over others, depending on the clinical situation (eg, in women with uncomplicated cystitis, fosfomycin trometamol is preferred over other antibiotics because it has the property “convenient protocol”). For each preference property, two junior doctors reviewed the icon web sites and designed personalized icons using Inkscape. The seven preference properties, as described by CPG experts, are shown along with the representative icons in Table 1.

Figure 1. Basic wireframe of e-commerce interfaces compartmentalized into sections.

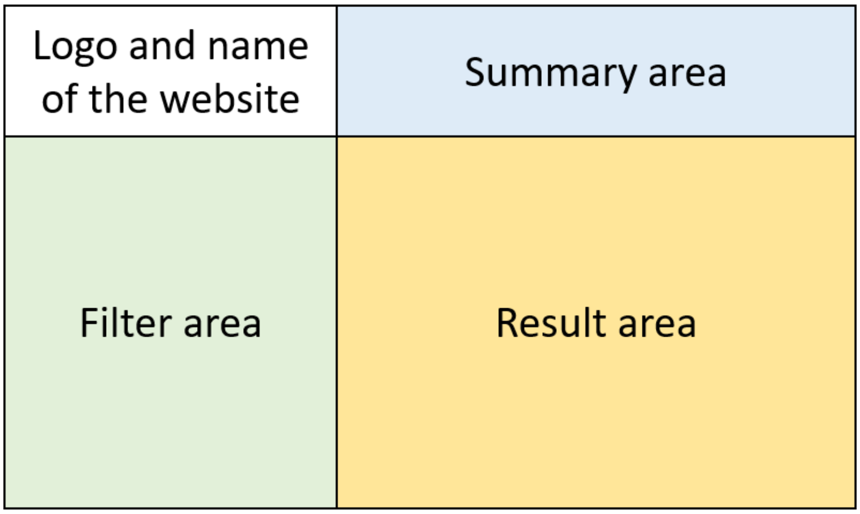


Table 1. Icons used for displaying antibiotic properties.

Preference property	Definition[a]	Icon description	Icon
Convenient protocol	If the antibiotic is prescribed orally and for less than Z days[b]	Symbolized by a capsule placed on a hand	
Absence of serious and frequent side effects	If there is no risk of serious side effects and the frequency of side effects is sufficiently low to allow prescription	Symbolized by a face with skin rash with an interdictory stroke	
Nonprecious class	If the antibiotic does not belong to a class of drugs that must be preserved for more serious infections	Symbolized by a diamond with an interdictory stroke	
Narrow antibacterial spectrum	If the antibiotic is described as having a narrow antibacterial spectrum	Symbolized by mass spectrum graph	
Low level of ecological adverse effects	If the antibiotic is described as having a low risk of promoting bacterial resistance	Symbolized by a recycling icon	
High level of efficacy	If the antibiotic is described as very effective (eg, high clinical cure percentage)	Symbolized by brachial biceps with a boxing glove	
Acceptable taste	If the antibiotic is described as having an acceptable taste for oral administration	Symbolized by a face with tongue emoji	

[a] Expert definitions within clinical practice guidelines. Source: [14].

[b] Z: the period depends on the clinical situation.

Interface Assessment

The utility of the interface for improving GPs’ comprehension of CPGs and their readability was evaluated using a questionnaire tailored to the study.

GP Recruitment

GPs were contacted via emails sent to medical networks in the Île-de-France region (including Paris) and by word of mouth, between March 25 and April 25, 2018. A reminder was sent 15 days after the first contact was made. For inclusion in the study, GPs had to be in training or practicing in primary care. The evaluation was voluntary and anonymous.

Study Design

A web-based evaluation was carried in three steps. First, the participating GPs were asked to answer five sociodemographic questions. Their responses were rendered anonymous. Next, the GPs were asked to use the web interface for the use case “young adult woman with uncomplicated pyelonephritis.” The information displayed on the web interface was derived from a knowledge base describing 11 infectious diseases, 50 antibiotics, and 21 patient profiles, constructed based on French CPGs and clinical expertise (a description of this knowledge base is available in the literature [14]). Then, the GPs were asked to answer 17 questions relating to usefulness (3 questions), user-friendliness (3 questions), satisfaction (3 questions) and comprehension (8 questions). The responses were recorded on a 4-point Likert scale (ie, “absolutely disagree,” “tend to disagree,” “tend to agree,” and “absolutely agree”). At the end of the evaluation, GPs were provided with an opportunity to write optional free comments.

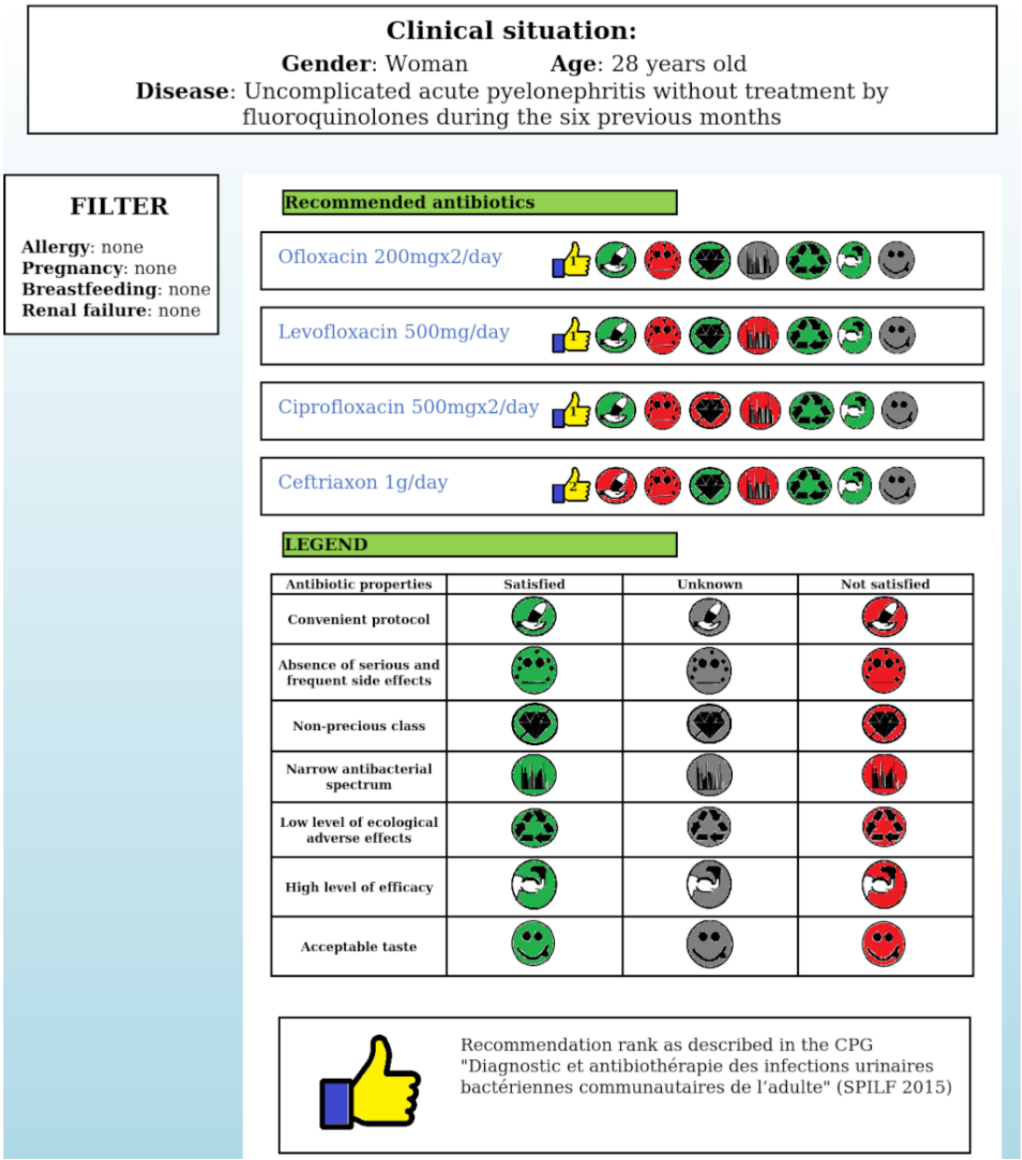
Results

Web Interface for Antibiotic Prescription

The web interface for antibiotic prescription was divided into three main sections, as in most e-commerce websites (Figure 2):

1. The *clinical summary* section, located at the top of the interface, displays a short summary of the clinical situation defined by both the disease and the profile of the patient (eg, sex, age).
2. The *filter* section, located at the top left of the interface, can be used to filter the recommended drugs according to patient-specific conditions such as drug allergies, renal failure, pregnancy, and/or breastfeeding.
3. The *recommended antibiotics* section, located in the center of the interface, displays the recommended drugs, with justifications in the form of drug properties. These preference properties are represented by seven icons. Each icon is shown in a different color, as follows: green, if the recommended antibiotic satisfies the property; red, if the recommended antibiotic does not satisfy the property; and grey, if no information is available. For each recommended drug, the recommendation rank is displayed with a numbered thumbs-up icon, as per the CPGs.
4. The *legend* located at the bottom of the interface highlights the icons used in the recommended antibiotics section and displays hypertextual links to the original CPGs.

Figure 2. Web interface for antibiotic prescription displaying recommended drugs and their properties.



Clinical situation:

Gender: Woman **Age**: 28 years old

Disease: Uncomplicated acute pyelonephritis without treatment by fluoroquinolones during the six previous months

FILTER

Allergy: none
Pregnancy: none
Breastfeeding: none
Renal failure: none

Recommended antibiotics

Ofloxacin 200mgx2/day 👍1 🟢 🔴 🟢 ⚪ 🟢 🟢 ⚪
Levofloxacin 500mg/day 👍1 🟢 🔴 🟢 🔴 🟢 🟢 ⚪
Ciprofloxacin 500mgx2/day 👍1 🟢 🔴 🔴 🔴 🟢 🟢 ⚪
Ceftriaxon 1g/day 👍2 🔴 🔴 🟢 🔴 🟢 🟢 ⚪

LEGEND

Antibiotic properties	Satisfied	Unknown	Not satisfied
Convenient protocol	🟢	⚪	🔴
Absence of serious and frequent side effects	🟢	⚪	🔴
Non-precious class	🟢	⚪	🔴
Narrow antibacterial spectrum	🟢	⚪	🔴
Low level of ecological adverse effects	🟢	⚪	🔴
High level of efficacy	🟢	⚪	🔴
Acceptable taste	🟢	⚪	🔴

👍 Recommendation rank as described in the CPG "Diagnostic et antibiothérapie des infections urinaires bactériennes communautaires de l’adulte" (SPILF 2015)

Evaluation of the Antibiotic Prescription Web Interface

Characteristics of Participating GPs

Given the opportunistic nature of the recruitment method used, it is difficult to estimate accurately how many GPs received the invitation email. The number of GPs contacted was estimated to range between 850 and 880.

In total, 117 GPs working within distinct GP surgeries accepted the invitation and assessed the web interface. More than half (73/117, 62.4%) were female. More than three-fourths (104/117, 89.0%) were under 40 years of age and had been in practice for less than 10 years. More than half (77/117, 65.8%) were working in private practice (Table 2).

Table 2. Sociodemographic characteristics of the general practitioners (GPs) (N=117).

Characteristic	Value, n (%)
Sex	
Female	73 (62.4)
Male	44 (37.6)
Age (years)	
20-30	42 (36.0)
30-40	62 (53.0)
40-50	7 (6.0)
50-60	3 (2.5)
>60	3 (2.5)
Professional status	
GP with an MD thesis	74 (63.3)
GP without an MD thesis	39 (33.3)
GP in training	4 (3.4)
Time in practice (years)	
<5	81 (69.2)
5-10	20 (17.1)
10-20	10 (8.5)
>20	6 (5.2)
Practice type	
Private	77 (65.8)
Salaried	15 (12.8)
Mixed	25 (21.4)

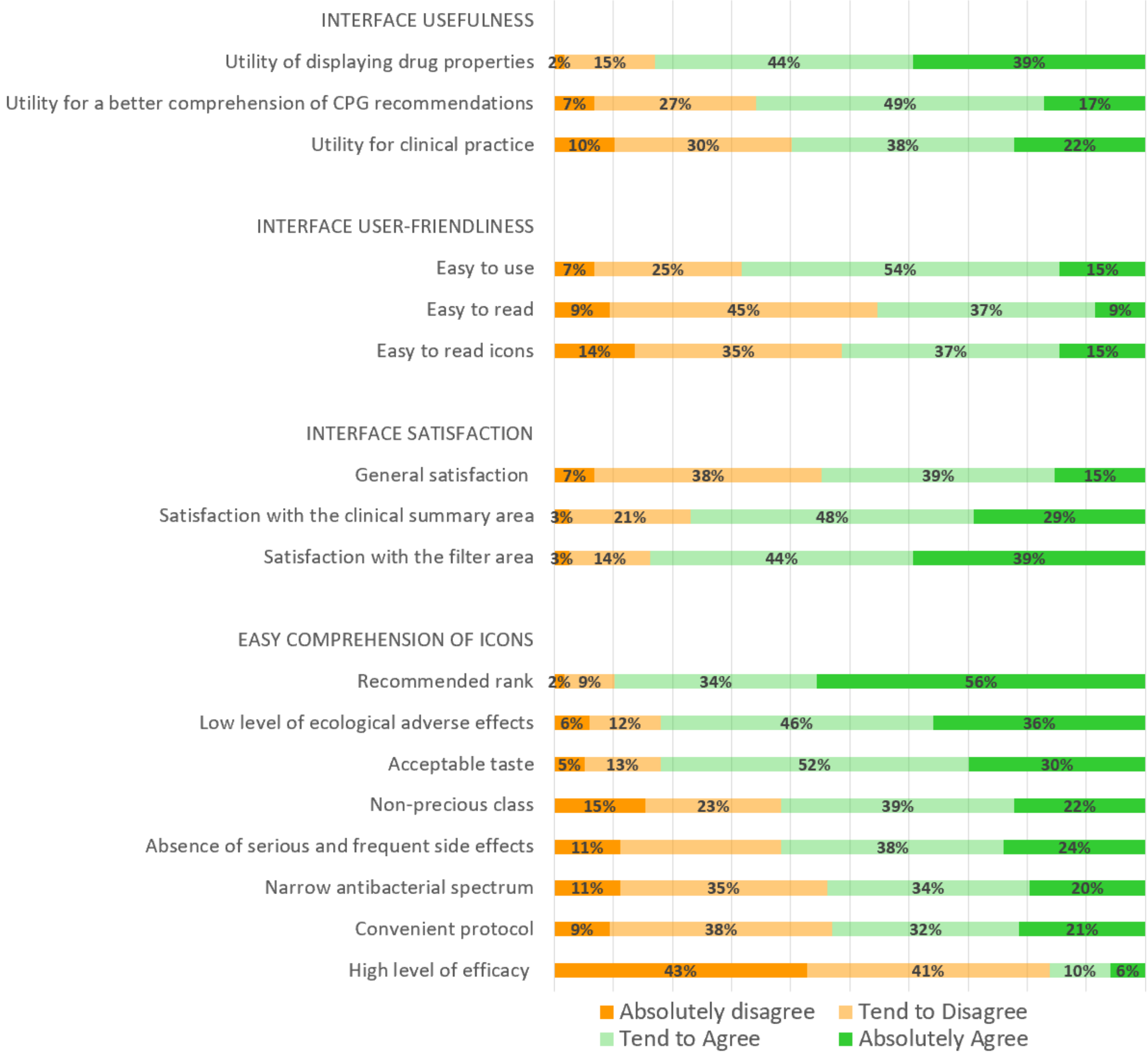
Satisfaction, Usefulness, Comprehensibility, and Readability of the Web Interface

Overall, 82.9% (97/117) of GPs found the display of drug properties useful (95% CI 76-90); 65.8% (77/117) reported that the interface improved their understanding of CPG recommendations (95% CI 57-74), and 59.8% (70/117) considered the interface useful for clinical practice (95% CI 51-69), as shown in Figure 3.

Only 54.7% (64/117) of all GPs were satisfied with the global interface (95% CI 46-64), probably because of dissatisfaction with some of the icons. Indeed, the majority of GPs appreciated the clinical summary (90/117, 76.9%) and filter (98/117, 83.8%) sections, whereas 48.7% (57/117) of them reported difficulties understanding some of the icons (Figure 3).

Five of the eight icons were considered sufficiently readable. The icons for recommended rank, ecological adverse effects, and taste were considered easily understandable by more than 80% of GPs; both the precious class and the side effects icons were considered comprehensible by more than 60% of all GPs. By contrast, three icons were not sufficiently readable; icons for antibacterial spectrum and convenient protocol were considered poorly understandable by nearly half the GPs and that for efficacy level was considered poorly understandable by about 84% of GPs (Figure 3).

Figure 3. Satisfaction, usefulness, readability, and comprehensibility of the web interface according to general practitioners. In the questionnaire, each item was turned into an affirmative sentence to be graded on a 4-point Likert scale ranging from “absolutely disagree” to “absolutely agree” (eg, the item “Utility of displaying drug properties” was converted to the statement “I found the display of drug properties useful” and this affirmation was then graded).



Areas for Improvement

Some GPs said that they found the web interface interesting but that it could take time for them to get familiar with it. They suggested several areas for improvement (Table 3).

For the clinical summary section, GPs suggested displaying more details about the probable causal microbes, patient history, treatment, and signs of gravity. For the filter section, they suggested adding filters for hepatic conditions and galenic forms. Regarding the recommended antibiotics section, the GPs thought it included too much information and needed to be simplified to improve readability. For example, some GPs said that only the most important properties should be displayed and that other properties should be hidden and accessible only on request (eg, nonprecious class, efficacy level, and taste). Surprisingly, some GPs reported that some properties were not important for the choice of antibiotic, whereas these properties were used by CPG experts in the formulation of recommendations (eg, ecological adverse effects, activity spectrum, antibiotic taste). Other GPs also had doubts about the nonprecious class property because they did not understand what this meant. Conversely, other GPs suggested including additional properties such as drug cost, drug-drug interactions, and contraindications. Some others said that some of the icons were not sufficiently intuitive and required revision. They suggested, for example, replacing the efficacy level by a graduated scale, speedometer, or weapon symbol. Finally, some GPs said that they thought the web interface could be useful for shared decision-making with patients.

Table 3. Areas for improvement extracted by analysis of free comments by general practitioners (GPs) (N=117).

Interface section and suggested improvement	GP, n (%)
Clinical summary	
Add more information about patient history and current treatment	9 (7.7)
Add more information about clinical signs of gravity	1 (0.9)
Add details about the suspected causal microbes	1 (0.9)
Filter	
Add hepatic condition filter	1 (0.9)
Add galenic form preference (eg, syrup, pill)	2 (1.7)
Recommended antibiotics	
Improve readability	31 (26.5)
Add some properties (eg, duration, cost)	8 (6.8)
Specify important patient contraindications for each antibiotic	2 (1.7)
Specify drug-drug interactions	3 (2.6)

Discussion

Principal Findings

An interface for antibiotic prescription displaying the recommendations and their justifications was designed and assessed. The web interface was divided into three main sections: a clinical summary section, a filter section, and a recommended antibiotics section enlisting recommended antibiotics and their properties displayed as icons. Overall, 82.9% of participating GPs found the display of drug properties useful, and 65.8% of them reported that the interface helped to improve their understanding of CPG recommendations. Nevertheless, the interface requires further improvement before its implementation in a clinical decision support system (CDSS).

Limitations

Interface Design

Icon Design

Despite 82.9% of the GPs reporting that they found the display of antibiotic properties useful, 48.7% reported difficulty in reading the corresponding icons. Starren et al [22] described five kinds of presentation for displaying medical data (ie, table, list, graph, generated text, and icon). We believed that icons were the most suitable presentation for displaying the antibiotic properties because they are small pictorial symbols that are particularly well adapted for displaying qualitative data within computer interfaces [22]. Furthermore, they are already widely used for displaying medical data such as for medication administration or patient events [22]. Iconic languages [23-25] have even been developed in the medical domain, such as the Visualisation des Connaissances Médicales (Visualization of Medical Knowledge [VCM]) language [25,26], which is used to represent signs, diseases, physiological states, life habits, and drugs. However, as there were no existing icons for representing antibiotic properties, the icons for this interface were designed by ourselves. Designing unambiguous icons for displaying precise concepts, such as “low level of ecological adverse effects” or “narrow antibacterial spectrum” was not an easy task, and it can take time for the GPs to become familiar with these icons. In our study, GPs discovered the icons for the first time during the evaluation period, with no prior tutorial or training, potentially accounting for their lack of satisfaction with the existing icons. The next step will be improving the design of some icons in accordance with the suggestions made by the GPs participating in this study (eg, “efficacy level” could be represented by a graduated scale). A tutorial and interactive information bubble will also be added to improve the readability of icons.

Missing Elements

The interface displayed the recommendations and the elements required for antibiotic prescription according to the experts writing CPGs. However, the qualitative analysis of the free comments highlighted the following: (1) some properties were considered important by CPG experts but not by GPs (eg, CPG experts considered antibiotic taste to be important for improving treatment adherence, whereas this was not the case for all GPs); (2) GPs considered other properties not mentioned in CPGs to be important (eg, some GPs reported needing to visualize drug interactions, which are seldom dealt with in CPGs); and (3) some properties are currently used by CPGs experts but are not understood by GPs (eg, the property nonprecious class). This gap between real-life practice and writing CPGs should be considered when designing CDSS interfaces. Thus, interfaces should also include the properties used by GPs in real-life practice, even if the evidence for their use is poor or not given in CPGs. For example, with regard to antibiotic prescription, Krishnakumar et al [6] developed a model of the rationale used by GPs for antibiotic choice. Some of the factors included in this model, such as drug pharmacokinetics, marketing authorization, and drug cost, will be included in the web interface in the future.

Interface Evaluation

The participating GPs were recruited via emails sent to medical networks and by word of mouth. This mode of recruitment was chosen because it is faster and cheaper than traditional approaches [27,28]. However, it may have introduced a selective participation bias [28] (eg, by selecting GPs keen on the use of new technologies). This bias was limited by using large mailing lists (>850 GPs), including GPs working in various health care centers, from a large French region (including Paris). The readability of the interface and its utility for improving GPs' comprehension of CPGs were also evaluated. Assessments of readability and comprehensibility are part of the software lifecycle and may prevent serious usability problems, which often occur during the design of new technologies [29]. Early testing makes it possible to identify serious issues and to improve the interface before its implementation in a CDSS.

Comparison With Previous Work

Comparison With Other Drug Comparison Interfaces

Other types of interfaces have been developed for comparing drugs in medical domains. With regard to Dopamine [30], for instance, double-entry tables were used to compare the contraindications and side effects of drugs. More sophisticated tables, including the possibilities of overlapping data and interactions, have also been developed, such as Twinlist [31] for medication reconciliation or rainbow boxes [32,33] for comparing drug information such as contraindications or side effects. Rainbow boxes were previously adapted for use in antibiotic prescription [34]; the resulting rainbow interface was perceived as easy to read for 27.5%-64% of GPs, depending on the clinical situation. However, as smartphones are increasingly being used by medical doctors for learning, information retrieval, and/or clinical decision support [35-38], it is important to consider interfaces more suitable for use with a smartphone. Tabular representations take up a large amount of space and are therefore more suitable for computer interfaces [39,40]. Here, an e-commerce interface adaptable for smartphones was designed as a means of overcoming this limitation. *Graphs* have also been used for drug comparisons, such as Rxplore [41] for the rapid review of potential drug events caused by drugs and Network graph [42] for chemotherapy treatment. However, in these cases, graphical representations were interesting because of the need to add quantitative information such as the results of clinical trials [42]. As antibiotic properties are qualitative data, graphical representations were not considered appropriate for this situation.

Comparison With Other Antibiotic Prescription Interfaces

Other interfaces have been developed for the empiric prescription of antibiotics in primary care, for example, IAAP Smart phone [43], ARI Smart Form [44], and ABX TRIP [21] for acute respiratory infections. These interfaces display antibiotic recommendations without their justifications—that is, only the recommended antibiotics are displayed in a textual format without their properties. This type of presentation suffers from "the black box effect" [45] and may make GPs passive in their decision-making process, simply accepting the decision suggested by the CDSS. GPs may also be suspicious of the recommendations due to a lack of understanding of the suggestions made by the CDSS. In our proposed interface, drug properties were displayed to involve GPs in the decision process. This improves their understanding of the underlying reasons for recommendations and enables them to compare the advantages and disadvantages of the recommended antibiotics. Preserving GP autonomy is an important factor to be considered when trying to increase the chances of CDSS being successfully adopted [46].

Conclusions

A web interface for antibiotic prescription presenting drug recommendations and their justifications was designed following the basic framework of an e-commerce interface. The justifications of CPG recommendations were displayed in the form of antibiotic properties, which was considered useful by more than three-fourths of GPs and helpful for understanding CPGs by two-thirds of GPs evaluating the interface. Further improvements are required before the implementation of this interface in a CDSS.

Acknowledgments

We thank the GPs for taking time to participate in this study despite their heavy workload. This study was funded by the Agence Nationale de Sécurité du Médicament et des Produits de santé (ANSM). APP 2016 – RaMiPA Project (Raisonner pour Mieux Prescrire les Antibiotiques or Reasoning for better antibiotic prescription) led by RT.

Authors' Contributions

RT led the project. RM and RT designed the study. RM, AU, DI, and RT designed the web interface.
RM and RT analyzed and interpreted the data. RT wrote the manuscript. All authors critically edited and approve the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Olesen SW, Barnett ML, MacFadden DR, Lipsitch M, Grad YH. Trends in outpatient antibiotic use and prescribing practice among US older adults, 2011-15: observational study. BMJ 2018 Jul 27;362:k3155 [FREE Full text] [doi: 10.1136/bmj.k3155] [Medline: 30054353]

2. van Houten CB, Cohen A, Engelhard D, Hays JP, Karlsson R, Moore E, et al. Antibiotic misuse in respiratory tract infections in children and adults-a prospective, multicentre study (TAILORED Treatment). Eur J Clin Microbiol Infect Dis 2019 Mar;38(3):505-514 [FREE Full text] [doi: 10.1007/s10096-018-03454-2] [Medline: 30707378]
3. Tamma PD, Avdic E, Li DX, Dzintars K, Cosgrove SE. Association of adverse events with antibiotic use in hospitalized patients. JAMA Intern Med 2017 Sep 01;177(9):1308-1315 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1938] [Medline: 28604925]
4. Marquet K, Liesenborgs A, Bergs J, Vleugels A, Claes N. Incidence and outcome of inappropriate in-hospital empiric antibiotics for severe infection: a systematic review and meta-analysis. Crit Care 2015 Feb 16;19:63 [FREE Full text] [doi: 10.1186/s13054-015-0795-y] [Medline: 25888181]
5. Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. P T 2015 Apr;40(4):277-283 [FREE Full text] [Medline: 25859123]
6. Krishnakumar J, Tsopra R. What rationale do GPs use to choose a particular antibiotic for a specific clinical situation? BMC Fam Pract 2019 Dec 20;20(1):178 [FREE Full text] [doi: 10.1186/s12875-019-1068-7] [Medline: 31862003]
7. Shively NR, Buehrle DJ, Clancy CJ, Decker BK. Prevalence of inappropriate antibiotic prescribing in primary care clinics within a Veterans Affairs health care system. Antimicrob Agents Chemother 2018 Aug;62(8):1-7. [doi: 10.1128/aac.00337-18]
8. Marc C, Vrignaud B, Levieux K, Robine A, Guen CG, Launay E. Inappropriate prescription of antibiotics in pediatric practice: analysis of the prescriptions in primary care. J Child Health Care 2016 Dec;20(4):530-536. [doi: 10.1177/1367493516643421] [Medline: 27091956]
9. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999 Oct 20;282(15):1458-1465. [doi: 10.1001/jama.282.15.1458] [Medline: 10535437]
10. Correa VC, Lugo-Agudelo LH, Aguirre-Acevedo DC, Contreras JAP, Borrero AMP, Patiño-Lugo DF, et al. Individual, health system, and contextual barriers and facilitators for the implementation of clinical practice guidelines: a systematic metareview. Health Res Policy Syst 2020 Jun 29;18(1):74 [FREE Full text] [doi: 10.1186/s12961-020-00588-8] [Medline: 32600417]
11. Bierbaum M, Rapport F, Arnolda G, Nic Giolla Easpaig B, Lamprell K, Hutchinson K, et al. Clinicians' attitudes and perceived barriers and facilitators to cancer treatment clinical practice guideline adherence: a systematic review of qualitative and quantitative literature. Implement Sci 2020 May 27;15(1):39 [FREE Full text] [doi: 10.1186/s13012-020-00991-3] [Medline: 32460797]
12. Tsopra R, Venot A, Duclos C. Towards evidence-based CDSSs implementing the medical reasoning contained in CPGs: application to antibiotic prescription. Stud Health Technol Inform 2014;205:13-17. [Medline: 25160136]
13. Tsopra R, Venot A, Duclos C. An algorithm using twelve properties of antibiotics to find the recommended antibiotics, as in CPGs. AMIA Annu Symp Proc 2014;2014:1115-1124 [FREE Full text] [Medline: 25954422]
14. Tsopra R, Lamy J, Sedki K. Using preference learning for detecting inconsistencies in clinical practice guidelines: Methods and application to antibiotherapy. Artif Intell Med 2018 Jul;89:24-33. [doi: 10.1016/j.artmed.2018.04.013] [Medline: 29776758]
15. Lamy J, Sedki K, Tsopra R. Explainable decision support through the learning and visualization of preferences from a formal ontology of antibiotic treatments. J Biomed Inform 2020 Apr;104:103407. [doi: 10.1016/j.jbi.2020.103407] [Medline: 32156641]
16. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
17. Miller K, Mosby D, Capan M, Kowalski R, Ratwani R, Noaiseh Y, et al. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. J Am Med Inform Assoc 2018 May 01;25(5):585-592. [doi: 10.1093/jamia/ocx118] [Medline: 29126196]
18. Shen Y, Yuan K, Chen D, Colloc J, Yang M, Li Y, et al. An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. Artif Intell Med 2018 Mar;86:20-32. [doi: 10.1016/j.artmed.2018.01.003] [Medline: 29433958]
19. Akhloufi H, Verhaegh SJC, Jaspers MWM, Melles DC, van der Sijs H, Verbon A. A usability study to improve a clinical decision support system for the prescription of antibiotic drugs. PLoS One 2019;14(9):e0223073 [FREE Full text] [doi: 10.1371/journal.pone.0223073] [Medline: 31553785]
20. Paul M, Andreassen S, Tacconelli E, Nielsen AD, Almanasreh N, Frank U, TREAT Study Group. Improving empirical antibiotic treatment using TREAT, a computerized decision support system: cluster randomized trial. J Antimicrob Chemother 2006 Dec;58(6):1238-1245. [doi: 10.1093/jac/dkl372] [Medline: 16998208]
21. Litvin CB, Ornstein SM, Wessell AM, Nemeth LS, Nietert PJ. Adoption of a clinical decision support system to promote judicious use of antibiotics for acute respiratory infections in primary care. Int J Med Inform 2012 Aug;81(8):521-526. [doi: 10.1016/j.ijmedinf.2012.03.002] [Medline: 22483528]
22. Starren J, Johnson SB. An object-oriented taxonomy of medical data presentations. J Am Med Inform Assoc 2000;7(1):1-20 [FREE Full text] [Medline: 10641959]

23. Preiss B, Kaltenbach M, Zanazaka J, Echave V. Concept graphics: a language for medical knowledge. Proc Annu Symp Comput Appl Med Care 1992:515-519 [FREE Full text] [Medline: 1482928]
24. Preiss B, Echavé V, Preiss SF, Kaltenbach M. UVAL-MED a universal visual associative language for medicine. Proc Annu Symp Comput Appl Med Care 1994:262-266 [FREE Full text] [Medline: 7949931]
25. Lamy J, Duclos C, Bar-Hen A, Ouvrard P, Venot A. An iconic language for the graphical representation of medical concepts. BMC Med Inform Decis Mak 2008 Apr 24;8:16 [FREE Full text] [doi: 10.1186/1472-6947-8-16] [Medline: 18435838]
26. Lamy J, Thuy VB, Lillo-Le Louët A, Bousquet C. An iconic approach to the browsing of medical terminologies. Stud Health Technol Inform 2019 Aug 21;264:213-217. [doi: 10.3233/SHTI190214] [Medline: 31437916]
27. Ali SH, Foreman J, Capasso A, Jones AM, Tozan Y, DiClemente RJ. Social media as a recruitment platform for a nationwide online survey of COVID-19 knowledge, beliefs, and practices in the United States: methodology and feasibility analysis. BMC Med Res Methodol 2020 May 13;20(1):1-11. [doi: 10.1186/s12874-020-01011-0]
28. Heiervang E, Goodman R. Advantages and limitations of web-based surveys: evidence from a child mental health survey. Soc Psychiatry Psychiatr Epidemiol 2011 Jan;46(1):69-76. [doi: 10.1007/s00127-009-0171-9] [Medline: 19921078]
29. Islam MN, Karim MM, Inan TT, Islam AKMN. Investigating usability of mobile health applications in Bangladesh. BMC Med Inform Decis Mak 2020 Feb 03;20(1):19 [FREE Full text] [doi: 10.1186/s12911-020-1033-3] [Medline: 32013965]
30. Wroe C, Solomon W, Rector A, Rogers J. DOPAMINE: a tool for visualizing clinical properties of generic drugs. 2000 Presented at: Proc Fifth Workshop Intell Data Anal Med Pharmacol IDAMAP; August 22, 2000; Germany p. 61-65.
31. Plaisant C, Wu J, Hettinger AZ, Powsner S, Shneiderman B. Novel user interface design for medication reconciliation: an evaluation of Twinlist. J Am Med Inform Assoc 2015 Mar;22(2):340-349. [doi: 10.1093/jamia/ocu021] [Medline: 25665706]
32. Lamy J, Berthelot H, Favre M, Ugon A, Duclos C, Venot A. Using visual analytics for presenting comparative information on new drugs. J Biomed Inform 2017 Jul;71:58-69 [FREE Full text] [doi: 10.1016/j.jbi.2017.04.019] [Medline: 28549568]
33. Lamy J, Tsopra R. RainBio: proportional visualization of large sets in Biology. IEEE Trans Vis Comput Graph 2020 Nov;26(11):3285-3298. [doi: 10.1109/TVCG.2019.2921544] [Medline: 31180862]
34. Tsopra R, Kinouani S, Venot A, Jaulent M, Duclos C, Lamy J. Design of a visual interface for comparing antibiotics using rainbow boxes. Stud Health Technol Inform 2017;235:529-533. [Medline: 28423849]
35. Nerminathan A, Harrison A, Phelps M, Alexander S, Scott KM. Doctors' use of mobile devices in the clinical setting: a mixed methods study. Intern Med J 2017 Dec;47(3):291-298. [doi: 10.1111/imj.13349] [Medline: 27925381]
36. Dimond R, Bullock A, Lovatt J, Stacey M. Mobile learning devices in the workplace: 'as much a part of the junior doctors' kit as a stethoscope'? BMC Med Educ 2016 Aug 17;16(1):207 [FREE Full text] [doi: 10.1186/s12909-016-0732-z] [Medline: 27530343]
37. Zhang C, Fan L, Chai Z, Yu C, Song J. Smartphone and medical application use among dentists in China. BMC Med Inform Decis Mak 2020 Sep 07;20(1):213 [FREE Full text] [doi: 10.1186/s12911-020-01238-3] [Medline: 32894112]
38. Maassen O, Fritsch S, Gantner J, Deffge S, Kunze J, Marx G, et al. Future mobile device usage, requirements, and expectations of physicians in German university hospitals: web-based survey. J Med Internet Res 2020 Dec 21;22(12):e23955 [FREE Full text] [doi: 10.2196/23955] [Medline: 33346735]
39. Tsopra R, Sedki K, Courtine M, Falcoff H, De Beco A, Madar R, et al. Helping GPs to extrapolate guideline recommendations to patients for whom there are no explicit recommendations, through the visualization of drug properties. The example of AntibioHelp® in bacterial diseases. J Am Med Inform Assoc 2019 Oct 01;26(10):1010-1019. [doi: 10.1093/jamia/ocz057] [Medline: 31077275]
40. Tsopra R, Jais JP, Venot A, Duclos C. Comparison of two kinds of interface, based on guided navigation or usability principles, for improving the adoption of computerized decision support systems: application to the prescription of antibiotics. J Am Med Inform Assoc 2014 Feb;21(e1):e107-e116 [FREE Full text] [doi: 10.1136/amiajnl-2013-002042] [Medline: 24008427]
41. Duke JD, Li X, Grannis SJ. Data visualization speeds review of potential adverse drug events in patients on multiple medications. J Biomed Inform 2010 Apr;43(2):326-331 [FREE Full text] [doi: 10.1016/j.jbi.2009.12.001] [Medline: 19995616]
42. Warner J, Yang P, Alterovitz G. Automated synthesis and visualization of a chemotherapy treatment regimen network. Stud Health Technol Inform 2013;192:62-66 [FREE Full text] [Medline: 23920516]
43. Smartphone application for antibiotic prescribing. National Centre for Infection Prevention and Management (CIPM) and Imperial College Healthcare NHS Trust. 2013 Jan 04. URL: https://webarchive.nationalarchives.gov.uk/20130104192657/http://www.dh.gov.uk/health/2012/08/smartphone-antibiotic-prescribing/ [accessed 2021-06-05]
44. Linder JA, Rose AF, Palchuk MB, Chang F, Schnipper JL, Chan JC, et al. Decision support for acute problems: the role of the standardized patient in usability testing. J Biomed Inform 2006 Dec;39(6):648-655 [FREE Full text] [doi: 10.1016/j.jbi.2005.12.002] [Medline: 16442853]
45. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019 Mar;28(3):231-237 [FREE Full text] [doi: 10.1136/bmjqs-2018-008370] [Medline: 30636200]
46. Esmaeilzadeh P, Sambasivan M, Kumar N, Nezakati H. Adoption of clinical decision support systems in a developing country: Antecedents and outcomes of physician's threat to perceived professional autonomy. Int J Med Inform 2015 Aug;84(8):548-560. [doi: 10.1016/j.ijmedinf.2015.03.007] [Medline: 25920928]

Abbreviations

CPG: clinical practice guideline
CDSS: clinical decision support system
GP: general practitioner
VCM: Visualisation des Connaissances Médicales (Visualization of Medical Knowledge)

Edited by R Kukafka; submitted 17.11.20; peer-reviewed by M Cuggia, E Andrikopoulou, C Alvarez, S Sabarguna, M Schmieding; comments to author 08.02.21; revised version received 24.02.21; accepted 27.04.21; published 11.06.21.

Please cite as:
Madar R, Ugon A, Ivanković D, Tsopra R
A Web Interface for Antibiotic Prescription Recommendations in Primary Care: User-Centered Design Approach
J Med Internet Res 2021;23(6):e25741
URL: https://www.jmir.org/2021/6/e25741
doi:10.2196/25741
PMID:34114958

Original Paper

Constructing High-Fidelity Phenotype Knowledge Graphs for Infectious Diseases With a Fine-Grained Semantic Information Model: Development and Usability Study

Lizong Deng[1,2*], PhD; Luming Chen[1,2*], MS; Tao Yang[1,2*], MM; Mi Liu[3], PhD; Shicheng Li[1,2], MM; Taijiao Jiang[1,2,4], PhD

[1]Center of Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]Suzhou Institute of Systems Medicine, Suzhou, China

[3]Jiangsu Institute of Clinical Immunology, Jiangsu Key Laboratory of Clinical Immunology, The First Affiliated Hospital of Soochow University, Suzhou, China

[4]Guangzhou Laboratory, Guangzhou, China

*these authors contributed equally

Corresponding Author:
Taijiao Jiang, PhD
Center of Systems Medicine
Institute of Basic Medical Sciences
Chinese Academy of Medical Sciences & Peking Union Medical College
#5 Dong Dan San Tiao
Dongcheng District
Beijing, 100005
China
Phone: 86 051262873781
Email: taijiao@ibms.pumc.edu.cn

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/7/e31481

Abstract

Background: Phenotypes characterize the clinical manifestations of diseases and provide important information for diagnosis. Therefore, the construction of phenotype knowledge graphs for diseases is valuable to the development of artificial intelligence in medicine. However, phenotype knowledge graphs in current knowledge bases such as WikiData and DBpedia are coarse-grained knowledge graphs because they only consider the core concepts of phenotypes while neglecting the details (attributes) associated with these phenotypes.

Objective: To characterize the details of disease phenotypes for clinical guidelines, we proposed a fine-grained semantic information model named PhenoSSU (semantic structured unit of phenotypes).

Methods: PhenoSSU is an "entity-attribute-value" model by its very nature, and it aims to capture the full semantic information underlying phenotype descriptions with a series of attributes and values. A total of 193 clinical guidelines for infectious diseases from Wikipedia were selected as the study corpus, and 12 attributes from SNOMED-CT were introduced into the PhenoSSU model based on the co-occurrences of phenotype concepts and attribute values. The expressive power of the PhenoSSU model was evaluated by analyzing whether PhenoSSU instances could capture the full semantics underlying the descriptions of the corresponding phenotypes. To automatically construct fine-grained phenotype knowledge graphs, a hybrid strategy that first recognized phenotype concepts with the MetaMap tool and then predicted the attribute values of phenotypes with machine learning classifiers was developed.

Results: Fine-grained phenotype knowledge graphs of 193 infectious diseases were manually constructed with the BRAT annotation tool. A total of 4020 PhenoSSU instances were annotated in these knowledge graphs, and 3757 of them (89.5%) were found to be able to capture the full semantics underlying the descriptions of the corresponding phenotypes listed in clinical guidelines. By comparison, other information models, such as the clinical element model and the HL7 fast health care interoperability resource model, could only capture the full semantics underlying 48.4% (2034/4020) and 21.8% (914/4020) of the descriptions of phenotypes listed in clinical guidelines, respectively. The hybrid strategy achieved an F1-score of 0.732 for the subtask of phenotype concept recognition and an average weighted accuracy of 0.776 for the subtask of attribute value prediction.

Conclusions: PhenoSSU is an effective information model for the precise representation of phenotype knowledge for clinical guidelines, and machine learning can be used to improve the efficiency of constructing PhenoSSU-based knowledge graphs. Our work will potentially shift the focus of medical knowledge engineering from a coarse-grained level to a more fine-grained level.

(J Med Internet Res 2021;23(6):e26892) doi:10.2196/26892

KEYWORDS

knowledge graph; knowledge granularity; machine learning; high-fidelity phenotyping; phenotyping; phenotype; semantic

Introduction

When people are sick, their bodies present a series of observable or perceptible abnormalities, which are called phenotypes. In medicine, the phenotype concept covers signs and symptoms, laboratory test results, and imaging findings [1]. Phenotypes characterize the clinical manifestations of diseases, which provide important clues for diagnoses. Knowledge about disease phenotypes is usually documented as free text in medical textbooks or clinical guidelines, and such knowledge forms are hard for computers to use. Therefore, it is essential to transform phenotype knowledge into a machine-understandable format to facilitate the development of automated systems that could improve health care [2].

To date, many structured knowledge bases, such as WikiData [3], MalaCards [4], and DBpedia [5], have been constructed for disease phenotypes. In these knowledge bases, the phenotype knowledge of a disease is represented as a list of phenotype concepts or terms (Multimedia Appendix 1). However, such a concept-based representation only focuses on the presence or absence of a phenotype but neglects its contextual properties [6,7]. The description "sudden, severe abdominal pain in the lower right abdomen," for example, names three attributes of abdominal pain, including the onset pattern (sudden), severity (severe), and quadrant pattern (lower right abdomen). These attributes are valuable for diagnosis but missing in the provided concept-based representation. Due to the neglect of phenotypic details, current phenotype knowledge bases only characterize disease manifestations at a very coarse-grained level [8], which is considered to be "sloppy and imprecise" [9,10].

To precisely represent phenotype knowledge in clinical guidelines, it is necessary to introduce fine-grained semantic information models [11], which consider phenotypes and attributes simultaneously. The currently available semantic models for representing phenotype information include but are not limited to clinical element models (CEMs) [12], the Health Level Seven fast health care interoperability resource (FHIR) model [13], and the clinical quality language model [14]. All these models can be viewed as standard entity-attribute-value structures, which represent phenotype information with sufficient details by using various attributes and qualifier values. For example, a CEM model considers 17 attributes associated with phenotypes, such as phenotype severity, laterality, and duration. Although semantic information models such as CEM and FHIR have defined many attributes for phenotypes, it should be noted that these attributes are mainly designed for recording phenotypic information in electronic medical records. However, the logic underlying phenotypic descriptions in clinical guidelines is quite different from that in electronic medical records. For example, clinical guidelines usually describe the frequency of a phenotype in a population (eg, 30% of patients may have a loss of smell); however, this attribute is not defined in CEM and FHIR models. In addition to the frequencies of phenotypes, other often used attributes such as temporal patterns (eg, acute, chronic) and pain characteristics (eg, sharp, dull) are also neglected by the CEM and FHIR models. Therefore, it is necessary to optimize the attributes included in current information models to make them more suitable for representing phenotype knowledge in clinical guidelines.

In this work, we aimed to develop a semantic information model that could effectively characterize the details of disease phenotypes for clinical guidelines. A semantic information model named PhenoSSU (semantic structured unit of phenotype) was developed based on the clinical guidelines for 193 infectious diseases from Wikipedia. A total of 12 attributes were included in PhenoSSU, which characterized the details of phenotypes from various aspects. Based on PhenoSSU, we constructed fine-grained phenotype knowledge graphs for these infectious diseases. Considering the increased annotation costs associated with the introduction of PhenoSSU, we also explored the potential of machine learning for performing automatic recognition for PhenoSSU based on free text. It is hoped that our work will contribute to the large-scale construction of fine-grained phenotype knowledge graphs for more diseases.

Methods

Materials

We collected the clinical guidelines for 193 infectious diseases from Wikipedia [15] as the corpus for constructing fine-grained phenotype knowledge graphs. In Wikipedia, the phenotypic knowledge of infectious diseases was usually buried in a section named signs and symptoms (Multimedia Appendix 1). Although Wikipedia is created and edited by volunteers worldwide, many studies have proven the high quality of its biomedical content [16,17]. In addition, phenotype knowledge graphs for WikiData [3] and DBpedia [5] were also constructed based on clinical guidelines from Wikipedia.

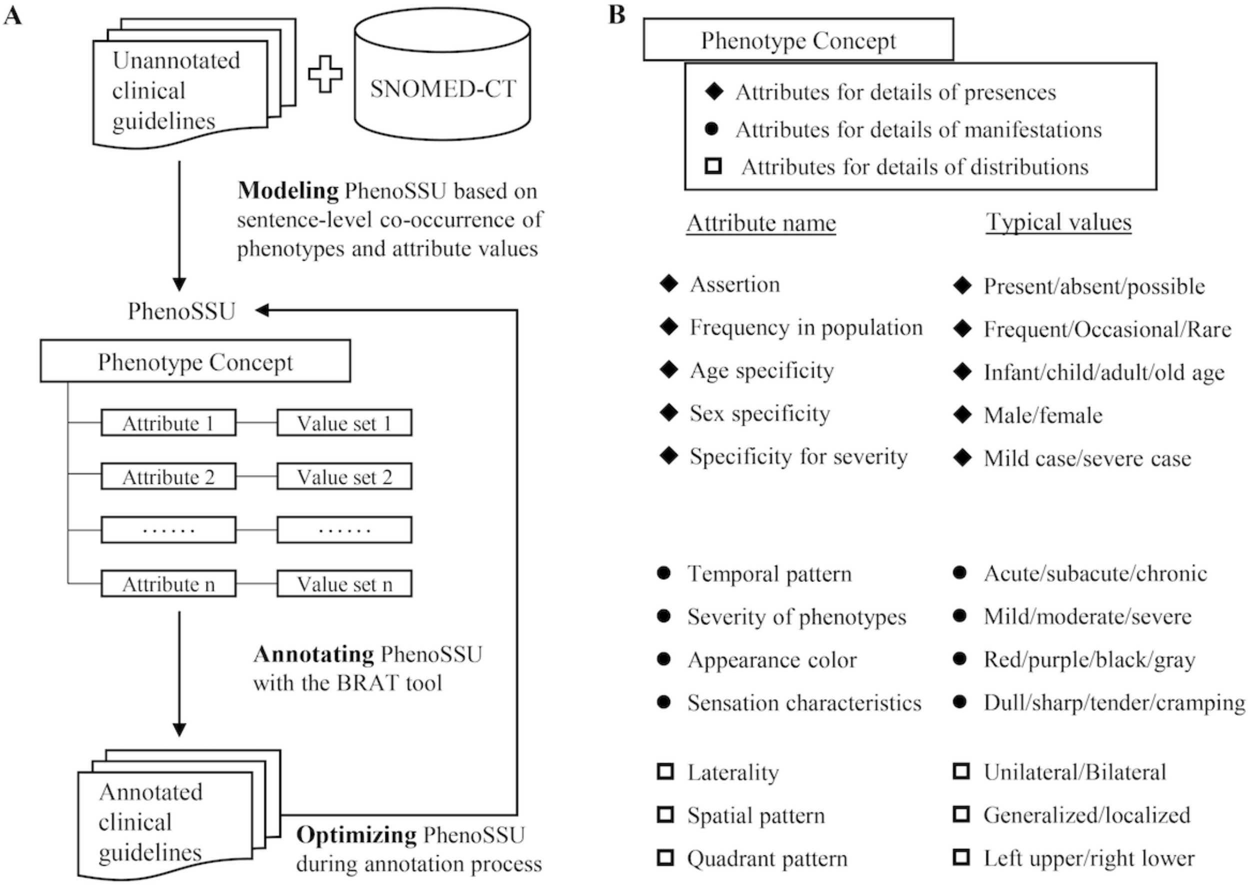
Design of PhenoSSU

PhenoSSU, by its very nature, is an entity-attribute-value model that consists of a phenotype concept along with a collection of attributes. Determining the attributes associated with various phenotypes is the key to the design of PhenoSSU. Four inclusion criteria for attributes were considered in this study:

- Introduced attribute and value set should come from a standard medical ontology to avoid the arbitrariness of defining new attributes. Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT) [18,19], one of the most comprehensive clinical terminology databases in the world, was selected as the standard for normalizing both phenotypes and attributes.
- Introduced attribute should be a modifier associated with phenotypes rather than an entity independent of phenotypes. The concepts found in SNOMED-CT were organized into 19 distinct hierarchies. Phenotypes and attributes were mainly located in the clinical finding and qualifier value hierarchies, respectively (Multimedia Appendix 1).
- Value set of the introduced attribute should contain categorical variables with limited dimensionality. For example, the severity attribute in SNOMED-CT contains a value set including mild, moderate, and severe. This criterion is for convenience when configuring attributes in the brat rapid annotation tool (BRAT) [20] (Multimedia Appendix 1).
- Introduced attribute should occur at least once in the studied corpus. This criterion is for reducing redundancy when introducing many unused attributes.

To effectively find the attributes associated with various phenotypes, we developed a simple co-occurrence–based method for attribute filtering (Figure 1A). Specifically, the phenotypes in the corpus were annotated with the MetaMap tool [21], a state-of-the-art concept recognizer, and the values of the attributes in the corpus were annotated with the Flashtext tool [22], a string-based concept recognizer. If an attribute co-occurred with any phenotypes in at least 2 sentences from the whole corpus, we selected the attribute as a candidate that was potentially associated with phenotypes. Then, we manually filtered the attributes that were truly related to phenotypes and built an initial version of PhenoSSU. The initial PhenoSSU model was optimized during the annotation process. When annotators found a new contextual property associated with phenotypes, we searched for its existence in SNOMED-CT and added the standard attribute corresponding to that contextual property into the initial PhenoSSU model.

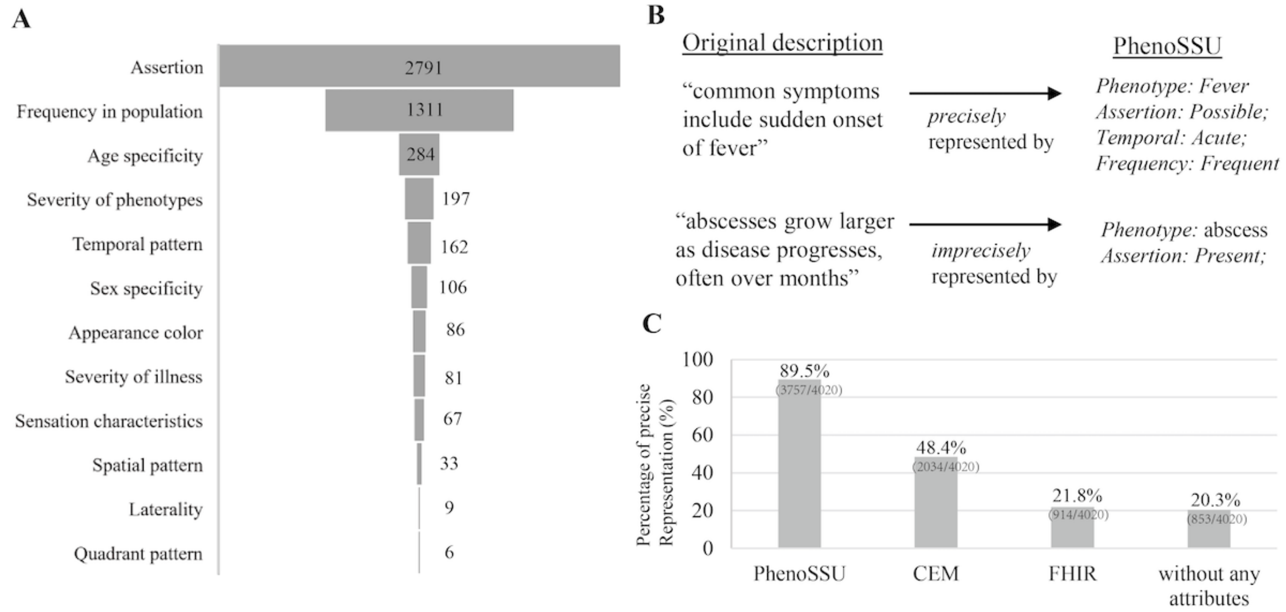
Figure 1. Modeling process of PhenoSSU: (A) modeling PhenoSSU based on sentence-level cooccurrences of phenotype concepts and attribute values in clinical guidelines and (B) components of the PhenoSSU model consist of a phenotype concept and 12 attributes.



The final PhenoSSU model contained 12 attributes, which could be classified into 3 categories according to the phenotypic details they characterized (Figure 1B): (1) details about the presence of phenotypes, including a phenotype's assertion, frequency in a population, age specificity, sex specificity, and specificity regarding the severity of illness; (2) details about the manifestations of phenotypes, including a phenotype's temporal pattern, severity, appearance color, and sensation characteristics;

and (3) details about the spatial distributions of phenotypes, including a phenotype's laterality, spatial pattern and quadrant pattern. The SNOMED-CT codes, definitions, and value sets of these attributes are listed in Multimedia Appendix 1. The distribution of these 12 attributes in the studied corpus is shown in Figure 2A.

Figure 2. Expressive power of PhenoSSU: (A) prevalence of the 12 attributes in the studied corpus, (B) examples of precise and imprecise representations for original phenotype descriptions with the PhenoSSU model, and (C) comparisons of precise representation percentages among different information models.



Annotation and Normalization of PhenoSSU

The annotation task of PhenoSSU can be divided into 2 steps: annotating a phenotype and annotating the attributes associated with that phenotype. Some annotation examples of different phenotypes attributes defined in PhenoSSU are presented in Multimedia Appendix 1. The clinical guides of 193 infectious diseases were annotated with the BRAT (Multimedia Appendix 1). To facilitate the annotation process, we preannotated the phenotypes found in clinical guidelines with the MetaMap tool. Then, two annotators (TY and SL) independently annotated the 193 clinical guidelines by following the annotation guide developed by LD and TJ. Their independent annotations were merged and visualized in the BRAT. To mark inconsistent annotations, we introduced a virtual attribute named agreement into PhenoSSU. Two independently annotated PhenoSSU models were regarded as consistent when both their phenotypes (text spans) and associated attribute values were the same. If there were inconsistencies in any part of a PhenoSSU model, the value of the agreement attribute was set to disagreement. The initial interannotator agreement at the PhenoSSU level was calculated with a Cohen kappa statistic [23] of 0.861. All inconsistent annotations were solved by an adjudication process (TJ).

The phenotypes annotated in BRAT were normalized with SNOMED-CT. To facilitate the normalization process, we also leveraged the MetaMap tool to obtain candidate concepts from the SNOMED-CT database and then manually selected the concept corresponding to each query phenotype. There was no need to normalize the attribute values because they were already normalized in SNOMED-CT.

One aspect to note about the normalization process is the special treatment used for finding sites of phenotypes. Finding sites were not explicitly included in the PhenoSSU model because they are entities independent of phenotypes. In SNOMED-CT, there were more than 39,000 concepts of finding sites in the body structure hierarchy, and these were hard to set as a value list in the BRAT. However, finding sites are indispensable information for describing phenotypes. Therefore, we also annotated the entities of finding sites associated with phenotypes. Taking the annotation of "bleeding from the nose and gum" as an example, the entities of the phenotype (bleeding) and two finding sites (nose, gum) were annotated separately and connected with a relation curve named locate (Multimedia Appendix 1). If a phenotype had an associated finding site, the phenotype together with the finding site was regarded as an integral concept in the normalization process. For example, the annotation of "bleeding" associated with "nose" was normalized as "249366005|epistaxis," which shared the same codes as the annotation of "bleeding from nose." If a composite concept could not be normalized as a whole (eg, "rash associated with hands"), we standardized the phenotype and its corresponding finding site separately and combined them into a postcoordination expression [24] (eg, "271807003|Rash": "33712006|Skin structure of hand"; Multimedia Appendix 1). In summary, information about finding sites was implicitly considered an integral part of a phenotype concept rather than its attribute.

Automatic Recognition of PhenoSSU

The manual annotation of a PhenoSSU model is a very time-consuming process because annotators not only need to find the mention of a phenotype but also need to determine the existence of attribute trigger terms in the context surrounding a phenotype. To reduce annotation costs, it is necessary to develop algorithms for the automatic annotation of PhenoSSU models.

The recognition task of PhenoSSU can be divided into 2 subtasks: phenotype concept recognition and attribute value prediction. The first subtask aims to recognize the text spans corresponding to phenotypes, and the second subtask aims to select appropriate values for 12 attributes based on a phenotype's context.

The 193 annotated clinical guides were randomly divided into a training set and a test set at a ratio of 6:4. For the subtask of phenotype concept recognition, we still used the MetaMap tool, which can recognize phenotype concepts based on the Metathesaurus in the Unified Medical Language System (2020AA release) [25]. We optimized the parameters of the MetaMap tool based on its performance on the task of recognizing phenotype concepts in the training set (Multimedia Appendix 1).

The subtask of attribute value prediction can be regarded as a classification problem, and two machine learning-based models were explored for this subtask. One model was based on a support vector machine (SVM), and the other model was based on a bidirectional long short-term memory (BiLSTM) neural network. For the value classification model of a specific attribute, the input was the encoded feature vectors of a phenotype's context and the output was one of the normalized values for this attribute.

We chose an SVM for developing attribute value prediction models because SVM-based models have proven their efficiency in the 2010 Informatics for Integrating Biology & the Bedside/Veterans Affairs challenge [26] and SemEval-2015 Task 14 [27]. In the SVM-based model (Multimedia Appendix 1), the context of a phenotype was encoded with the existence of trigger terms (terms that indicated a normalized value [eg, "sudden onset" was the trigger term of the normalized value "acute"]) and their distances to the target phenotype [26,27]. The SVM-based model was developed by using the scikit-learn package (version 0.23.1) [28]. The parameters of the SVM-based model were optimized by using a grid search strategy [29] on the training set.

Inspired by recent methodology developments for the assertion status prediction task [30,31], we chose BiLSTM for developing attribute value prediction models. The referenced studies [30,31] showed that BiLSTM and attention mechanisms could achieve better performances than other approaches when classifying assertions of medical concepts. Since assertion status prediction belonged to the task of attribute value prediction, we transferred the attention-enhanced BiLSTM model to our study. In a given BiLSTM-based model (Multimedia Appendix 1), the context of a phenotype was first split into 3 segments, including the left context, the phenotype itself, and the right context, which were then encoded into a 3×768 vector with a pretrained language model named BERT (bidirectional encoder representation from transformers) [32-34]. Each BiLSTM-based model was developed by using the Keras package (version 2.3.1) [35], and the BERT encoding process was performed by using the bert-as-service package (version 1.10.0) [36]. Considering the very imbalanced distributions of attribute values in our dataset (Multimedia Appendix 1), we used the synthetic minority oversampling technique [37] from the imbalanced-learn package (version 0.7.0) [38] to balance the sample distributions. The hyperparameters of the constructed BiLSTM-based models were optimized using an early stopping strategy [39] on the training set.

Evaluation of the Performance for Recognizing PhenoSSU

To evaluate the performance of the proposed algorithm in extracting PhenoSSU models from free text, we used the evaluation metrics from SemEval-2015 Task 14: Analysis of Clinical Text [27].

The evaluation metric for the subtask of phenotype concept recognition was the F1-score. A predicted phenotype concept was regarded as a true positive if its text span overlapped with a gold standard text span. The precision metric was calculated as the fraction of correctly predicted phenotypes among all phenotypes identified by MetaMap, and the recall metric was calculated as the fraction of correctly predicted phenotypes among all phenotypes identified by the annotators. The F1-score was calculated as the harmonic mean of precision and recall.

We chose the average weighted accuracy as the evaluation metric for the subtask of attribute value prediction because the distributions of different attribute values were very imbalanced. The average weighted accuracy metric considers the prevalence of an attribute value in the corpus, so it can measure how good an algorithm is at predicting the rare values of an attribute. The detailed calculating process of the average weighted accuracy can be found in Multimedia Appendix 1.

Evaluation of the Expressive Power of PhenoSSU

Since the aim of this work was to develop a semantic information model that was more suitable than current approaches for representing phenotype knowledge in clinical guidelines, it was necessary to evaluate whether the annotated PhenoSSU model could capture the full semantics underlying the original descriptions of phenotypes. For example, in Figure 2B, the description "common symptoms include sudden onset of fever" could be perfectly represented by the PhenoSSU model (phenotype: fever; assertion: possible; frequency: frequent; temporal pattern: acute). By comparison, the description "abscesses grow larger as disease progress, often over months" was only partially represented by the PhenoSSU model (phenotype: abscess; assertion: present), which missed the information regarding the course and duration of abscess associated with the description.

To evaluate the expressive power of PhenoSSU, we introduced a virtual attribute named "equal to the original description" into the PhenoSSU model. If the annotated PhenoSSU did not capture the full semantics of an original description, we set the value of this attribute to "partial." Two annotators (TY and SL) independently evaluated the expressive power of the annotated PhenoSSU model. The initial interannotator agreement as measured with Cohen kappa statistic was 0.903 (3631/4020).

We reached a consensus for those inconsistent judgments by an adjudication process (TJ).

Results

Overview of the PhenoSSU Model and PhenoSSU-Based Knowledge Graphs

To characterize the details of phenotypes for clinical guidelines, a semantic information model named PhenoSSU was proposed. With the introduction of 12 attributes associated with various phenotypes, the obtained knowledge graphs based on PhenoSSU were more fine-grained than those based on phenotype concepts. In this work, 193 PhenoSSU-based knowledge graphs for infectious diseases were constructed. At the concept level, we annotated 4020 phenotypic terms, 3962 of which could be normalized with 1508 concepts in SNOMED-CT. At the attribute level, we annotated 5278 nondefault attribute values ("present" was the default attribute value for the assertion attribute, and "none" was the default attribute value for other attributes), which indicated the widespread presence of contextual properties for phenotypes in clinical guides. The most commonly used attributes included assertion, frequency in a population, age specificity, phenotype severity, and temporal pattern (Figure 2A).

Since the knowledge graphs in WikiData were also extracted from Wikipedia, we compared our knowledge graphs with those in WikiData at the concept level. WikiData built knowledge graphs for 66 of the 193 diseases, and these graphs included 354 phenotype concepts. Our annotations covered 297 of the 354 (83.9%) phenotypes from WikiData. For the uncovered phenotypes, we could not confirm their existence on the corresponding webpages of Wikipedia (including current and historical webpages). Most of these uncovered phenotypes may come from the manual additions of volunteers, who made use of sources other than Wikipedia (Multimedia Appendix 1).

Expressive Power of PhenoSSU for Representing Phenotype Knowledge

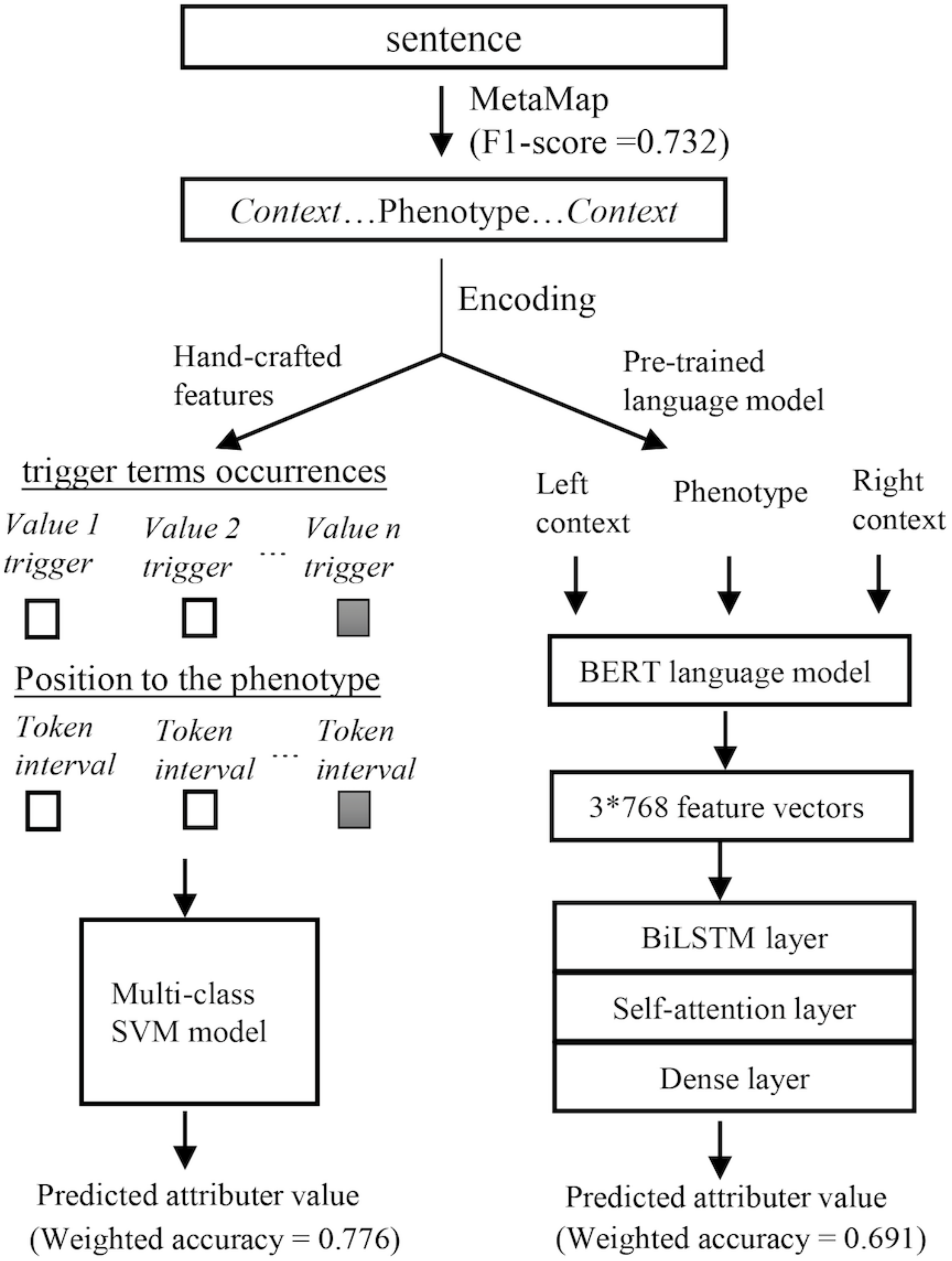
To evaluate the expressive power of the PhenoSSU model quantitatively, we manually analyzed whether a PhenoSSU instance could capture the full semantics underlying the corresponding descriptions of phenotypes (Figure 2B).

In this study, we annotated 4020 PhenoSSU instances, 3757 of which (89.5%) were determined to precisely represent the original phenotype knowledge described by natural language (Figure 2C). If we only considered the presence and absence of phenotype concepts (concept-based representation), the percentage of precise representations decreased to 20.3% (853/4200). This result further suggested the necessity of introducing the attributes associated with phenotypes into the developed model. We also analyzed the expressive power of the CEM and FHIR models for phenotypes and found that their percentages of precise representations were 48.4% (2034/4200) and 21.8% (914/4200), respectively. Most of the attributes defined in the CEM and FHIR models were not used in clinical guidelines except for the severity and laterality of phenotypes. The CEM model achieved a higher expression power than that of the FHIR model because it considered the uncertainty of phenotypes (assertion: possible), which is a frequently used attribute in clinical guidelines. Please see Multimedia Appendix 1 for detailed comparisons between the attributes used in the PhenoSSU, CEM, and FHIR models.

Potential for Increasing the Speed of PhenoSSU Model Annotation With Machine Learning

With the introduction of attributes, it would take more time to annotate a PhenoSSU model than to annotate phenotype concepts. To increase the efficiency of annotating PhenoSSU models, we developed a hybrid strategy that first recognized phenotype concepts with the MetaMap tool and then predicted the attribute values of phenotypes with SVM-based or BiLSTM-based classifiers (Figure 3). For the subtask of phenotype concept recognition, the MetaMap tool achieved an F1-score of 0.732 (precision 0.660; recall 0.824), which was comparable to its performance on other medical corpora [40]. For the subtask of attribute value prediction, the average weighted accuracy of the SVM-based method (0.776) was better than that of the BiLSTM-based model (0.691). This may be due to limited number of training data, which made it hard for the deep learning-based approach to learn useful features from contexts. However, the performance of the BiLSTM-based model was still higher than the performance of a reference model (0.542) that always selected default values for attributes (it selected "present" for the assertion attribute and "none" for other attributes). These results indicate that machine learning methods have the potential to speed up PhenoSSU annotations. The detailed performances of the compared models for predicting the values of different attributes are listed in Multimedia Appendix 1.

Figure 3. Automatic recognition of PhenoSSU.



Discussion

Principal Findings

In this work, we designed a fine-grained information model named PhenoSSU, which can precisely represent phenotype knowledge for clinical guidelines. We also developed an automatic strategy to extract PhenoSSU models from clinical guidelines and found that machine learning could be used to improve the efficiency of PhenoSSU annotation. Taken together, our work will provide a useful theoretical and technical guide for the construction of fine-grained phenotype knowledge graphs.

From the design of PhenoSSU, it can be seen that PhenoSSU was derived from SNOMED-CT because both the phenotype concepts and attribute values in PhenoSSU came from SNOMED-CT. PhenoSSU strengthened the expressive power of SNOMED-CT by combining 12 attributes with phenotype concepts. In SNOMED-CT, there was a technique named postcoordination expression [24] that could also capture the details of phenotypes by using combinations of existing concepts. For example, the out-of-vocabulary concept “severe headache, unilateral” can be expressed as a postcoordination of 3 concepts—headache (25064002): severity (272141005) = severe (24484000) and laterality (272741003) = unilateral (66459002). Compared with the postcoordination expression technique, PhenoSSU is a predefined information model that provides a general framework for knowledge representation. It is more convenient to configure the PhenoSSU model into the BRAT annotation tool to construct fine-grained phenotype knowledge graphs than to use the competing approach.

In recent years, machine learning, especially deep learning, has been widely used for processing medical information [41-44]. In this work, we also explored the potential of automatically constructing fine-grained phenotype knowledge graphs based on machine learning. The results in Figure 3 suggest that machine learning can assist with the human annotations of PhenoSSU to some extent. However, there are still great challenges to overcome to improve the performance of machine learning, especially the insufficiency and imbalanced distributions of training data. In future work, an active learning framework [45] that incorporates both human intelligence and machine intelligence may be a better strategy for constructing fine-grained knowledge graphs.

The improvement of knowledge granularity for disease phenotypes may potentially benefit knowledge-based diagnosis systems because the differential diagnostic capability of a PhenoSSU model is theoretically stronger than that of a single phenotype concept. From the perspective of coarse-grained knowledge graphs, some diseases (eg, the flu and common cold) have many similar symptoms (eg, fever and cough); however, these similar symptoms may have obvious differences from the perspectives of fine-grained knowledge graphs. For example, fever may be present in both flu and common cold. However, fever is more common in flu patients and usually appears suddenly with a body temperature of 38 degrees or above. By comparison, fever is rarely seen in common cold cases and usually appears gradually. Therefore, a diagnosis system cannot exclude the common cold if a patient has fever; however, it can safely exclude the common cold if a patient has such a PhenoSSU instance like “phenotype: fever; temporal pattern: acute; severity: severe.” PhenoSSU-based knowledge graphs should be very suitable for dialogue-based symptom checkers such as babylon [46] and symptoma [47], which collects the symptoms of a patient one by one. Considering the details of phenotypes in inquiry processing may potentially improve the efficiency and accuracy of dialogue-based symptom checkers.

Limitations

One limitation of this work is that we only considered the corpus of infectious diseases during the modeling process of PhenoSSU. In addition, we only considered attributes with categorical values and did not consider attributes with numeric values. Another limitation of this study is that we only tested the effectiveness of the PhenoSSU model for 193 infectious diseases, which is a small number considering that thousands of other diseases exist. In addition, attributes suitable for infectious diseases may not be suitable for other types of diseases. We will solve these limitations during the process of constructing PhenoSSU-based knowledge graphs for more diseases in future work.

The annotation guidelines for PhenoSSU and the PhenoSSU-based knowledge graphs for 193 infectious diseases can be found by visiting our website [48]. The scripts for modeling and extracting PhenoSSU can be found on GitHub [49].

Conclusions

PhenoSSU is a fine-grained semantic information model that can precisely represent phenotype knowledge in clinical guidelines, and machine learning can be used to improve the efficiency of constructing PhenoSSU-based knowledge graphs.

Acknowledgments

This work was supported by grants 32070678 and 31671371 from the National Natural Science Foundation of China, and grant EKPG21-12 from Emergency Key Program of Guangzhou Laboratory， and grants 2016-I2M-1-005 and 2020-I2M-2-003 from the Chinese Academy of Medical Sciences Initiative for Innovative Medicine. We sincerely thank colleagues in our lab and experts in the biomedical field for their thoughtful suggestions to improve this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures, tables and texts.
[PDF File (Adobe PDF File), 933 KB - jmir_v23i6e26892_app1.pdf]

References

1. Brookes AJ, Robinson PN. Human genotype-phenotype databases: aims, challenges and opportunities. Nat Rev Genet 2015 Dec;16(12):702-715. [doi: 10.1038/nrg3932] [Medline: 26553330]
2. Oellrich A, Collier N, Groza T, Rebholz-Schuhmann D, Shah N, Bodenreider O, et al. The digital revolution in phenotyping. Brief Bioinform 2016 Sep;17(5):819-830 [FREE Full text] [doi: 10.1093/bib/bbv083] [Medline: 26420780]
3. Turki H, Shafee T, Hadj Taieb MA, Ben Aouicha M, Vrandečić D, Das D, et al. Wikidata: a large-scale collaborative ontological medical database. J Biomed Inform 2019 Nov;99:103292. [doi: 10.1016/j.jbi.2019.103292] [Medline: 31557529]

4. Rappaport N, Twik M, Plaschkes I, Nudel R, Iny Stein T, Levitt J, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. Nucleic Acids Res 2017 Jan 04;45(D1):D877-D887 [FREE Full text] [doi: 10.1093/nar/gkw1012] [Medline: 27899610]
5. Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes P. DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. Semantic web 2015;6(2):167-169. [doi: 10.3233/sw-140134]
6. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. N Engl J Med 2018 Dec 11;379(15):1452-1462. [doi: 10.1056/NEJMra1615014] [Medline: 30304648]
7. Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. J Biomed Inform 2009 Oct;42(5):839-851 [FREE Full text] [doi: 10.1016/j.jbi.2009.05.002] [Medline: 19435614]
8. Hripcsak G, Albers DJ. High-fidelity phenotyping: richness and freedom from bias. J Am Med Inform Assoc 2018 Mar 01;25(3):289-294 [FREE Full text] [doi: 10.1093/jamia/ocx110] [Medline: 29040596]
9. Robinson PN. Deep phenotyping for precision medicine. Hum Mutat 2012 May;33(5):777-780. [doi: 10.1002/humu.22080] [Medline: 22504886]
10. Delude CM. Deep phenotyping: the details of disease. Nature 2015 Nov 4;527(7576):S14-S15. [doi: 10.1038/527s14a]
11. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
12. Oniki TA, Zhuo N, Beebe CE, Liu H, Coyle JF, Parker CG, et al. Clinical element models in the SHARPn consortium. J Am Med Inform Assoc 2016 Mar;23(2):248-256 [FREE Full text] [doi: 10.1093/jamia/ocv134] [Medline: 26568604]
13. Bender D, Sartipi K. HL7 FHIR: an agile and RESTful approach to healthcare information exchange. 2013 Presented at: Proc 26th IEEE Int Symp Computer-Based Med Syst; 2013; Porto. [doi: 10.1109/cbms.2013.6627810]
14. Clinical Quality Language. URL: https://cql.hl7.org/ [accessed 2021-03-30]
15. Wikipedia list of infectious diseases. URL: https://en.wikipedia.org/wiki/List_of_infectious_diseases [accessed 2020-09-30]
16. Rajagopalan MS, Khanna VK, Leiter Y, Stott M, Showalter TN, Dicker AP, et al. Patient-oriented cancer information on the internet: a comparison of wikipedia and a professionally maintained database. J Oncol Pract 2011 Sep;7(5):319-323 [FREE Full text] [doi: 10.1200/JOP.2010.000209] [Medline: 22211130]
17. Reavley NJ, Mackinnon AJ, Morgan AJ, Alvarez-Jimenez M, Hetrick SE, Killackey E, et al. Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. Psychol Med 2012 Aug;42(8):1753-1762. [doi: 10.1017/S003329171100287X] [Medline: 22166182]
18. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Stud Health Technol Inform 2006;121:279-290. [Medline: 17095826]
19. Gaudet-Blavignac C, Foufi V, Bjelogrlic M, Lovis C. Use of the systematized nomenclature of medicine clinical terms (SNOMED CT) for processing free text in health care: systematic scoping review. J Med Internet Res 2021 Jan 26;23(1):e24594 [FREE Full text] [doi: 10.2196/24594] [Medline: 33496673]
20. Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S. BRAT: a web-based tool for NLP-assisted text annotation. 2012 Presented at: The 13th Conference of the European Chapter of the Association for Computational Linguistics; April 23-27 2012; Avignon p. 102-107 URL: https://www.aclweb.org/anthology/E12-2021.pdf
21. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
22. Singh V. Replace or retrieve keywords in documents at scale. ArXiv. Preprint posted online on November 9, 2017 [FREE Full text]
23. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
24. Dhombres F, Winnenburg R, Case J, Bodenreider O. Extending the coverage of phenotypes in SNOMED CT through post-coordination. Stud Health Technol Inform 2015;216:795. [doi: 10.1007/978-1-84882-803-2_12] [Medline: 26262161]
25. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
26. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011;18(5):552-556 [FREE Full text] [doi: 10.1136/amiajnl-2011-000203] [Medline: 21685143]
27. Elhadad N, Pradhan S, Gorman S, Manandhar S, Chapman W, Savova G. SemEval-2015 task 14: analysis of clinical text. 2015 Presented at: The 9th International Workshop on Semantic Evaluation (SemEval 2015); June 4-5, 2015; Denver p. 303-310. [doi: 10.18653/v1/s15-2051]
28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: machine learning in Python. J Machine Learn Res 2011;12:2825-2830. [doi: 10.5555/1953048.2078195]
29. Tuning the hyper-parameters of an estimator. URL: https://scikit-learn.org/stable/modules/grid_search.html [accessed 2020-10-20]
30. Zhang ZC, Zhang Y, Zhou T, Pang YL. Medical assertion classification in Chinese EMRs using attention enhanced neural network. Math Biosci Eng 2019 Mar 08;16(4):1966-1977 [FREE Full text] [doi: 10.3934/mbe.2019096] [Medline: 31137195]

31. Kocaman V, Talby D. Improving clinical document understanding on COVID-19 research with spark NLP. ArXiv. Preprint posted online on December 7, 2020 [FREE Full text]
32. Devlin J, Chang M, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. Arxiv. Preprint posted online on May 24, 2019 [FREE Full text]
33. Luo Y, Cheng Y, Uzuner, Szolovits P, Starren J. Segment convolutional neural networks (Seg-CNNs) for classifying relations in clinical notes. J Am Med Inform Assoc 2018 Jan 01;25(1):93-98 [FREE Full text] [doi: 10.1093/jamia/ocx090] [Medline: 29025149]
34. Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H. Attention-based bidirectional long short-term memory networks for relation classification. 2016 Presented at: Proc 54th Annu Mtg Assoc Comp Linguist; 2016; Berlin p. 207-212. [doi: 10.18653/v1/p16-2034]
35. Gulli A, Pal S. Deep Learning with Keras. Birmingham: Packt Publishing Ltd; 2017.
36. Xiao H. bert-as-service. URL: https://github.com/hanxiao/bert-as-service [accessed 2020-09-20]
37. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
38. Lemaître G, Nogueira F, Aridas C. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. J Machine Learn Res 2017;18(1):559-563 [FREE Full text]
39. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep Learning. Cambridge: MIT Press; 2016.
40. Demner-Fushman D, Rogers WJ, Aronson AR. MetaMap Lite: an evaluation of a new Java implementation of MetaMap. J Am Med Inform Assoc 2017 Jul 01;24(4):841-844 [FREE Full text] [doi: 10.1093/jamia/ocw177] [Medline: 28130331]
41. Wu S, Roberts K, Datta S, Du J, Ji Z, Si Y, et al. Deep learning in clinical natural language processing: a methodical review. J Am Med Inform Assoc 2020 Mar 01;27(3):457-470. [doi: 10.1093/jamia/ocz200] [Medline: 31794016]
42. Wang J, Deng H, Liu B, Hu A, Liang J, Fan L, et al. Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: bibliometric study on PubMed. J Med Internet Res 2020 Jan 23;22(1):e16816 [FREE Full text] [doi: 10.2196/16816] [Medline: 32012074]
43. Li R, Yin C, Yang S, Qian B, Zhang P. Marrying medical domain knowledge with deep learning on electronic health records: a deep visual analytics approach. J Med Internet Res 2020 Sep 28;22(9):e20645 [FREE Full text] [doi: 10.2196/20645] [Medline: 32985996]
44. Liu H, Zhang Z, Xu Y, Wang N, Huang Y, Yang Z, et al. Use of BERT (Bidirectional Encoder Representations from Transformers)-based deep learning method for extracting evidences in chinese radiology reports: development of a computer-aided liver cancer diagnosis framework. J Med Internet Res 2021 Jan 12;23(1):e19689 [FREE Full text] [doi: 10.2196/19689] [Medline: 33433395]
45. Chen Y, Carroll RJ, Hinz ERM, Shah A, Eyler AE, Denny JC, et al. Applying active learning to high-throughput phenotyping algorithms for electronic health records data. J Am Med Inform Assoc 2013 Dec;20(e2):e253-e259 [FREE Full text] [doi: 10.1136/amiajnl-2013-001945] [Medline: 23851443]
46. Ćirković A. Evaluation of four artificial intelligence-assisted self-diagnosis apps on three diagnoses: two-year follow-up study. J Med Internet Res 2020 Dec 04;22(12):e18097 [FREE Full text] [doi: 10.2196/18097] [Medline: 33275113]
47. Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, et al. Diagnostic accuracy of web-based COVID-19 symptom checkers: comparison study. J Med Internet Res 2020 Oct 06;22(10):e21299 [FREE Full text] [doi: 10.2196/21299] [Medline: 33001828]
48. Database of PhenoSSU. URL: https://www.jianglab.tech/PhenoSSU/ [accessed 2021-03-30]
49. Deng L. Scripts for PhenoSSU. URL: https://github.com/denglizong/scripts-for-PhenoSSU [accessed 2021-03-30]

Abbreviations

BERT: bidirectional encoder representation from transformers
BiLSTM: bidirectional long short-term memory
BRAT: brat rapid annotation tool
CEM: clinical element model
FHIR: fast health care interoperability resource
PhenoSSU: semantic structured unit of phenotype
SNOMED-CT: Systematized Nomenclature of Medicine–Clinical Terms
SVM: support vector machine

Edited by R Kukafka; submitted 02.01.21; peer-reviewed by J Pacheco, C Seebode, JA Benítez-Andrades; comments to author 20.03.21; revised version received 01.04.21; accepted 06.05.21; published 15.06.21.

Please cite as:
Deng L, Chen L, Yang T, Liu M, Li S, Jiang T
Constructing High-Fidelity Phenotype Knowledge Graphs for Infectious Diseases With a Fine-Grained Semantic Information Model: Development and Usability Study
J Med Internet Res 2021;23(6):e26892
URL: https://www.jmir.org/2021/6/e26892
doi:10.2196/26892
PMID:34128811

Original Paper

Use of Multiprognostic Index Domain Scores, Clinical Data, and Machine Learning to Improve 12-Month Mortality Risk Prediction in Older Hospitalized Patients: Prospective Cohort Study

Richard John Woodman[1], BSc, MMedSci, MBiostat, PhD; Kimberley Bryant[1], BSc, PhD; Michael J Sorich[1], BSc, PhD; Alberto Pilotto[2,3], MD; Arduino Aleksander Mangoni[1], MD, PhD

[1]College of Medicine and Public Health, Flinders University, Adelaide, Australia
[2]Department of Geriatric Care, OrthoGeriatrics and Rehabilitation, Galliera Hospital, Genoa, Italy
[3]Department of Interdisciplinary Medicine, University of Bari, Bari, Italy

Corresponding Author:
Richard John Woodman, BSc, MMedSci, MBiostat, PhD
College of Medicine and Public Health
Flinders University
Room 3,12 Health Sciences Building
Sturt Road, Bedford Park
Adelaide, 5042
Australia
Phone: 61 0872218537
Email: richard.woodman@flinders.edu.au

Abstract

Background: The Multidimensional Prognostic Index (MPI) is an aggregate, comprehensive, geriatric assessment scoring system derived from eight domains that predict adverse outcomes, including 12-month mortality. However, the prediction accuracy of using the three MPI categories (mild, moderate, and severe risk) was relatively poor in a study of older hospitalized Australian patients. Prediction modeling using the component domains of the MPI together with additional clinical features and machine learning (ML) algorithms might improve prediction accuracy.

Objective: This study aims to assess whether the accuracy of prediction for 12-month mortality using logistic regression with maximum likelihood estimation (LR-MLE) with the 3-category MPI together with age and gender (feature set 1) can be improved with the addition of 10 clinical features (sodium, hemoglobin, albumin, creatinine, urea, urea-to-creatinine ratio, estimated glomerular filtration rate, C-reactive protein, BMI, and anticholinergic risk score; feature set 2) and the replacement of the 3-category MPI in feature sets 1 and 2 with the eight separate MPI domains (feature sets 3 and 4, respectively), and to assess the prediction accuracy of the ML algorithms using the same feature sets.

Methods: MPI and clinical features were collected from patients aged 65 years and above who were admitted to either the general medical or acute care of the elderly wards of a South Australian hospital between September 2015 and February 2017. The diagnostic accuracy of LR-MLE was assessed together with nine ML algorithms: decision trees, random forests, extreme gradient boosting (XGBoost), support-vector machines, naïve Bayes, K-nearest neighbors, ridge regression, logistic regression without regularization, and neural networks. A 70:30 training set:test set split of the data and a grid search of hyper-parameters with 10-fold cross-validation—was used during model training. The area under the curve was used as the primary measure of accuracy.

Results: A total of 737 patients (female: 370/737, 50.2%; male: 367/737, 49.8%) with a median age of 80 (IQR 72-86) years had complete MPI data recorded on admission and had completed the 12-month follow-up. The area under the receiver operating curve for LR-MLE was 0.632, 0.688, 0.738, and 0.757 for feature sets 1 to 4, respectively. The best overall accuracy for the nine ML algorithms was obtained using the XGBoost algorithm (0.635, 0.706, 0.756, and 0.757 for feature sets 1 to 4, respectively).

Conclusions: The use of MPI domains with LR-MLE considerably improved the prediction accuracy compared with that obtained using the traditional 3-category MPI. The XGBoost ML algorithm slightly improved accuracy compared with LR-MLE, and adding clinical data improved accuracy. These results build on previous work on the MPI and suggest that implementing risk scores based on MPI domains and clinical data by using ML prediction models can support clinical decision-making with respect to risk stratification for the follow-up care of older hospitalized patients.

(J Med Internet Res 2021;23(6):e26139) doi:10.2196/26139

KEYWORDS

machine learning; Multidimensional Prognostic Index; mortality; diagnostic accuracy; XGBoost

Introduction

Background

Previous studies have highlighted the importance of using functional measures to predict mortality among older hospitalized patients, a complex population characterized by different degrees of frailty, comorbidity burden, and polypharmacy [1,2]. In particular, the Multidimensional Prognostic Index (MPI), an objective and quantifiable comprehensive geriatric assessment (CGA)–based tool developed from 8 separate CGA domains, is an example of a cumulative deficit model of frailty [3]. The MPI is strongly associated with mortality [2], length of hospital stay [4], and hospital readmission [3] in subpopulations suffering from acute and chronic conditions, including general and geriatric hospital patients. The MPI patient score is created by aggregating the total scores from the 8 separate CGA tools and normalizing the resulting total score to provide a value in the range from 0 to 1. The latter is then categorized into three categories of risk: low (0.0-0.33), moderate (0.34-0.66), and severe (0.67-1.0), allowing clinicians to better tailor their care management.

Prediction accuracy for 12-month mortality (12MM) using the area under the receiver operating curve (AUC) of 0.751 was achieved with the three MPI categories when validated in an older Italian population [2]. However, despite it being a significant predictor of 12MM in a similar older hospital population within Australia [5], its diagnostic accuracy was not confirmed (AUC=0.64, with age and gender adjustment). The relatively poor performance for prediction compared with the prediction accuracy with the Italian cohort might be partly explained by the homogenization of scores from the separate MPI domains into a single aggregate-weighted scoring system. Although this simplifies risk classification, the use of an aggregate-weighted scoring system has been shown in general to remove important domain-specific information, resulting in poorer risk prediction [6]. Conversely, the use of individual MPI domains in place of an aggregate score has the potential to cause overfitting of the prediction model (ie, reduced bias), which results in lower accuracy when used on independent data sets (ie, increased variance) [7].

Machine learning (ML) is a branch of artificial intelligence in which various algorithms are used to make predictions [8]. The algorithms differ from standard statistical modeling approaches such as those using least squares or maximum likelihood, which focus on linear relationships and have no additional aspects to their respective error functions (such as the use of regularization) that help reduce the likelihood of model overfitting. The strengths of ML algorithms include their ability to handle feature selection in the presence of collinearity, and the ability to deal with a larger number of features including complex nonlinear patterns and interactions [9]. Furthermore, the validation of ML-based approaches is generally more rigorous than that of the standard statistical approaches, with special care taken to consider the aforementioned trade-off between bias and variance when developing the prediction model during the training process [10]. Finally, model development is more data driven; it does not rely solely on content knowledge, thereby increasing the opportunity to identify previously unconsidered features for enhanced prediction [8].

Objective

Given the relatively poor performance of the aggregate MPI score in predicting 12MM in an Australian cohort, we seek to improve the prediction accuracy in several ways. First, we use the separate components of the MPI as input features for a traditional logistic regression with maximum likelihood estimation (LR-MLE). Second, we assess nine different binary-classification ML algorithms that might perform better than LR-MLE. Third, we add 10 routinely collected clinical measures to the MPI domain-based feature data set.

Methods

Overview

Details of the data collection methods, including a description of the study design, study cohort, and collected data, have been previously published [5]. Briefly, the cohort consisted of patients aged 65 years and above admitted to the Flinders Medical Centre Acute Medical Unit and then transferred to either the general medical or acute care of the elderly wards between September 14, 2015, and February 17, 2017. Flinders Medical Center is a 593-bed metropolitan teaching and trauma hospital within the Southern Adelaide Local Health Network, which has a catchment area of approximately 350,000 people. Acute care of the elderly wards provides a comprehensive individualized approach for assessing older frail medical inpatients using a multidisciplinary team. The study was conducted in accordance with the Declaration of Helsinki and the guidelines for Good Clinical Practice. Approval for the study was obtained from the local ethics committee (reference number: 170.15).

Feature Sets

The 63-item MPI is a prognostic tool based on averaging the standardized scores obtained from the eight core domains of the CGA, which were obtained in all study participants within the first 3 days of hospital admission. We trained the prediction models using four different feature sets. The first feature set contained age, gender, and the 3-category MPI as three separate dummy variables (n=5 features in total). The second feature set contained the five features from the first feature set as well as 10 additional clinical features: BMI, anticholinergic risk score (ARS) [11], serum sodium, hemoglobin, serum albumin, creatinine, urea, urea-to-creatinine ratio, estimated glomerular filtration rate (eGFR), and C-reactive protein (CRP; n=15 features in total). The third and fourth feature sets included the eight separate MPI domains used in the calculation of the overall MPI score, in place of the 3-category MPI used in the first and second feature sets, resulting in n=10 and n=20 features, respectively. The eight MPI domains consist of cohabitation status (living alone, with family or friends, or in an institute), the total number of prescribed medications (taken at admission), functional status evaluated with activities of daily living (ADL) and instrumental ADL (IADL) scales; cognitive status evaluated by the Short Portable Mental Status Questionnaire; evaluation of pressure sores using the Exton Smith Scale (ESS); comorbidities assessed using the Cumulative Illness Rating Scale (CIRS); and nutritional status evaluated by the Mini Nutritional Assessment (MNA). The additional biochemical features are known to be associated with adverse outcomes in older patient populations [12-18]. The target variable for the prediction models was all-cause mortality within 12 months (12MM), defined as death from any cause and obtained using the Australian national death registry.

ML Algorithms

We implemented a systematic ML-based framework to construct the 12MM prediction models. The steps included data preprocessing, splitting of the data into training and validation data sets, model development using the training data set for each algorithm, and final assessment of the accuracy of each algorithm using the validation data set. The data preprocessing step included imputation of missing values and the scaling of continuous features, which included all features except for gender, the 3-category MPI, and cohabitation status. Any continuous features with missing values were imputed using the mean value of that feature before scaling the features and data splitting. Continuous features were scaled to have a zero mean and unit variance. Following preprocessing, the data were split into a training data set for the development of the prediction models and a test data set for accuracy validation. Data were split randomly into two sets in the ratio 70:30 with the 69.9% (515/737) sample defined as the training set and used for development of the prediction models, and the 30.1% (222/737) sample defined as the test (ie, validation) set and used to validate the accuracy of the algorithm. Once the training set was defined, an optimal model was developed for nine different ML algorithms: decision trees (DTs), random forests (RFs), eXtreme Gradient Boosting (XGBoost), support vector machines (SVM), naïve Bayes, K-nearest neighbors (KNN), ridge regression (logistic regression with *L2* regularization), logistic regression without regularization, and neural networks (NNs). A description of each algorithm is provided in Multimedia Appendix 1. Multimedia Appendix 2 provides the data set for the study and Multimedia Appendices 3-7 contain the Python code for all of the analysis performed for this study. All features were used without creating interactions or higher-order terms. However, as a sensitivity analysis, we also assessed each algorithm using all linear and second-order polynomial terms, that is, after squaring each feature and including all 2-way interactions.

For each algorithm, a grid search of hyper-parameters was performed to find the optimal set of hyper-parameters for training data accuracy. Each grid search was performed using 10-fold cross-validation, in which the training data set was split into 10 equally sized discrete folds. A model was then created using 90% (9/10 folds) of the data, and its accuracy was assessed using the remaining fold of data. The process was repeated 10 times, with each fold held out for one of the 10 training steps and used to assess the model accuracy for the training data. The AUC was used as the accuracy metric during the grid search. Once the optimal set of hyper-parameters was defined for each of the algorithms based on the training data, the performance of the optimal model for each ML algorithm was assessed on the test data set using the AUC as the primary accuracy metric. Although numerous accuracy measures are available for ML algorithms, the AUC is the most used in clinical settings and allows comparison with other studies, and we therefore used this as our primary accuracy measure. However, for completeness, we also report accuracy, precision, recall, and F1 score, given the imbalance in the number of positive and negative outcomes (dead and alive patients) that can, by itself, lead to higher AUC values [19].

Logistic Regression With Maximum Likelihood Estimation

For each feature set, the estimated logit coefficients obtained for the LR-MLE using the training data set were used to predict and assess the accuracy of the model on the test data set. The odds ratios (ORs), 95% CIs, and *P* values for the training data set with feature set 4 are reported.

Statistical Analysis

All analyses were performed using Python version 3.8.3. The normal distribution of the features was assessed using quantile-quantile plots and histograms, and descriptive statistics, including the mean, median, or frequency, were used for each feature as appropriate. Between-group comparisons for those alive and deceased at 12 months were performed using two-tailed independent *t* tests, Mann-Whitney tests, or chi-square tests. Each ML algorithm was implemented using Python's scikit-learn library [20], except for the XGBoost algorithm, which has its own Python package [21]. Relative importance feature plots and calibration plots were produced for the best algorithm, and violin plots were used to describe the distribution of the most important features. The calibration of the best-performing algorithm was shown by plotting the observed versus predicted deciles of risk. LR-MLE was performed using the logit function of Python's *statsmodels* module, and the LR-MLE models included all features within each feature set. Descriptive statistics were analyzed using the SciPy library (version 1.4.1) *stats* module, and plots were drawn using the *matplotlib* and *seaborn* libraries.

Data Sharing Statement

All data generated or analyzed during this study are included in Multimedia Appendix 2. The Python code used to analyze the data is also available in the Multimedia Appendices 3-7.

Results

Overview

The cohort included a total of 737 patients that were each assessed for MPI and followed up for 12 months. There were no missing values for age, sex, or each of the MPI domains. Among the additional five variables used in feature set 4, there were a total of 66 missing values, including sodium (n=1) albumin (n=3) hemoglobin (n=2), urea (n=1) creatinine (n=1), urea-to-creatinine ratio (n=1), eGFR (n=1), and CRP (n=56). Table 1 describes the characteristics of the patients according to their vital status at 12 months after mean imputation for missing values. There were significant differences between the two groups for all 20 features except for sex, the total number of medications used, the Short Portable Mental Status Questionnaire score, eGFR, creatinine, and the ARS.

Table 1. Patient characteristics according to vital status at 12 months after hospital discharge.

Characteristics	Alive (n=536)	Deceased (n=201)	*P* value[a]
Age (years), median (IQR)	79 (72-85)	82 (74-88)	.002
Gender, n (%)			.14
Female	278 (51.9)	92 (45.6)	
Male	257 (47.9)	110 (54.4)	
MPI[b] category, n (%)			<.001
Mild	211 (39.4)	39 (19.3)	
Moderate	290 (54.2)	136 (67.3)	
Severe	34 (6.4)	27 (13.4)	
MPI domains			
ADL[c], median (IQR)	6 (5-6)	5 (4-6)	<.001
IADL[d], median (IQR)	6 (4-8)	4 (3-6)	<.001
SPMSQ[e], median (IQR)	1 (0-2)	1 (0-3)	.05
ESS[f], median (IQR)	18 (17-19)	17 (15-18)	<.001
CIRS[g], mean (SD)	2.4 (0.4)	2.6 (0.4)	<.001
MNA[h], mean (SD)	20.9 (3.9)	18.0 (4.6)	<.001
Total number of medications, mean (SD)	10.0 (4.4)	10.3 (4.5)	.55
Cohabitation status, n (%)			.009
Living alone	199 (37.2)	70 (34.7)	
Family or friends	300 (56.1)	104 (51.5)	
Institute	36 (6.7)	28 (13.9)	
BMI (kg/m^2), median (IQR)	26.9 (23.8-31.8)	25.2 (22.0-29.1)	<.001
Sodium (mmol/L), median (IQR)	138 (135-140)	138 (135-140)	.006
Albumin (g/L), mean (SD)	32.5 (5.6)	30.4 (5.7)	<.001
Hemoglobin (g/L), mean (SD)	118.7 (18.2)	112.0 (18.3)	<.001
eGFR[i] (mL/min/1.73m^2), mean (SD)	55.3 (24.2)	52.2 (26.6)	.14
CRP[j] (mg/L), median (IQR)	29.0 (6.0-81.0)	33.0 (14.2-84.0)	.048
Creatinine (mmol/L), median (IQR)	95 (74-134)	103 (72-151)	.09
Urea (mmol/L), median (IQR)	7.40 (5.3-11.6)	8.90 (5.7-15.0)	<.001
Urea-to-creatinine ratio, median (IQR)	0.08 (0.06-0.10)	0.09 (0.06-0.10)	.001
ARS[k], median (IQR)	0 (0-2)	0 (0-2)	.41

[a] Using two-tailed independent *t* test, Mann-Whitney U test, or chi-square test, as appropriate.

[b] MPI: Multidimensional Prognostic Index.

[c] ADL: activities of daily living.

[d] IADL: instrumental activities of daily living.

[e] SPMSQ: Short Portable Mental Status Questionnaire.

[f] ESS: Exton Smith Scale.

[g] CIRS: Cumulative Illness Rating Scale.

[h] MNA: Mini Nutritional Assessment.

[i] eGFR: estimated glomerular filtration rate.

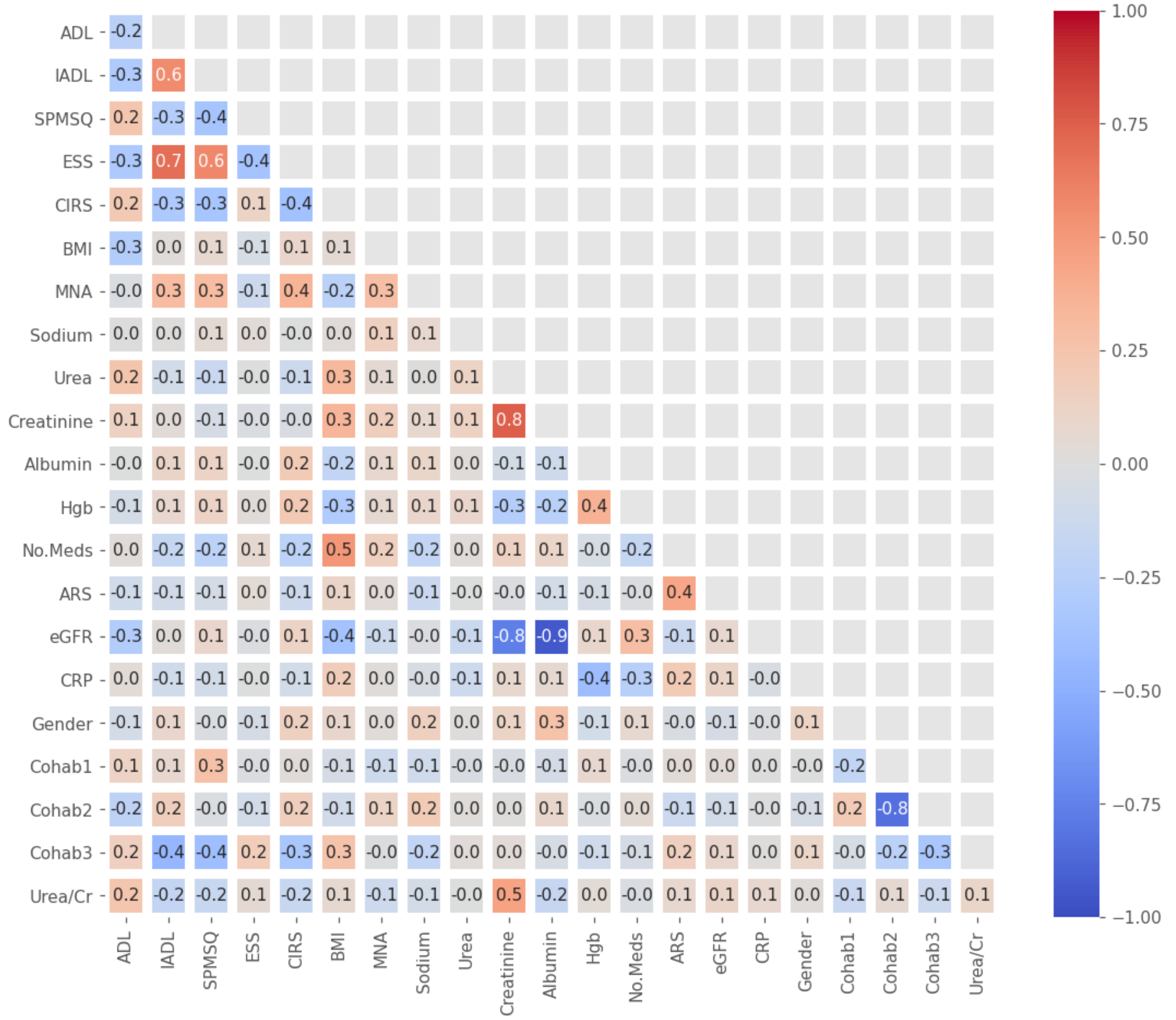
[j] CRP: C-reactive protein.

[k] ARS: anticholinergic risk score.

Correlation Matrix Heatmap

Figure 1 shows the Spearman ρ correlation matrix heatmap for features in feature set 4. Moderate to strong positive correlations were observed between creatinine and urea (ρ=0.75), ADL and ESS (ρ=0.69), IADL and ESS (ρ=0.58), ADL and IADL (ρ=0.57), and between CIRS and number of medications (ρ=0.51). There were also strong to moderate negative correlations between cohabitation status 1 and 2, that is, living alone and living with family or friends (ρ=−0.84), eGFR and creatinine (ρ=−0.94), eGFR and urea (ρ=−0.77), and CRP and albumin (ρ=−0.41). The absolute strengths of all other correlations were ｜ρ｜≤0.40. The lack of many highly correlated features suggested that the use of data reduction techniques such as principal component analysis before modeling was unnecessary.

Figure 1. Spearman ρ correlation matrix heatmap for feature set 4. ADL: activities of daily living; ARS: anticholinergic risk score; CIRS: Cumulative Illness Rating Scale; Cohab1: living alone; Cohab2: living with family or friends; Cohab3: living in an institute; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; ESS: Exton Smith Scale; Hgb: serum hemoglobin; IADL: instrumental activities of daily living; MNA: Mini Nutritional Assessment; No.Meds: number of medications; SPMSQ: Short Portable Mental Status Questionnaire; Urea/Cr: urea-to-creatinine ratio.



ML Algorithms

Test Data Accuracy

Table 2 and Figure 2 describe the test accuracy results for the four feature sets. The AUC for LR-MLE for feature sets 3 and 4 (0.738 and 0.757, respectively) that contained the eight MPI domains were considerably higher than those for feature sets 1 and 2 (0.632 and 0.688, respectively) that contained the three MPI categories. The AUC for LR-MLE was lower than those for at least one of the ML algorithms for each feature set 1 to 3 and was very similar to the best ML algorithms for feature set 4 (0.757 for LR-MLE vs 0.757 for XGBoost and 0.758 for NN). Overall, the best-performing ML algorithm was XGBoost, with an AUC ranging from 0.635 to 0.757 for feature sets 1 and 4, respectively, and a mean AUC of 0.714 for all four feature sets. The AUC for LR-MLE and all nine ML algorithms was improved with the addition of the clinical data (feature set 2 vs feature set 1 and feature set 4 vs feature set 3). The AUC was also improved for some, but not all, of the ML algorithms with the addition of clinical data. Multimedia Appendix 8 provides the results of accuracy, precision, recall, and F1 score for the LR-MLR model and the nine ML algorithms. As with the AUC, values for accuracy and precision were comparable across the various models with a range of 0.707 (KNN) to 0.752 (NN) for accuracy and a range of 0.474 (KNN) to 0.654 (SVM) for precision. However, there was a wider variability for recall and F1 scores with recall ranging from 0.031 (DT) to 0.531 (naïve Bayes) and the F1-score ranging from 0.059 (DT) to 0.751 (SVM).

Table 2. Diagnostic accuracy for logistic regression with maximum likelihood estimation and the 9 machine learning algorithms using feature sets 1 to 4 with the test data set.

Model	AUC[a]				
	Feature set 1[b]	Feature set 2[c,d]	Feature set 3[e]	Feature set 4[f,d]	Value, mean (SD)
LR-MLE[g]	0.632	0.688	0.738	0.757	0.704 (0.06)
Machine learning algorithms					
XGB[h]	0.635	0.706	0.756	0.757	0.714 (0.06)
Neural network	0.637	0.689	0.749	0.758	0.708 (0.06)
Random forest	0.621	0.684	0.753	0.751	0.702 (0.06)
Ridge[i]	0.632	0.671	0.738	0.749	0.698 (0.06)
KNN[j]	0.626	0.642	0.731	0.715	0.679 (0.06)
Nonpenalized logistic regression	0.627	0.642	0.707	0.690	0.667 (0.05)
Naïve Bayes	0.591	0.649	0.705	0.704	0.663 (0.04)
SVM[k]	0.530	0.661	0.737	0.711	0.656 (0.09)
Decision tree	0.604	0.588	0.695	0.686	0.643 (0.06)

[a] AUC: area under the receiver operating curve.

[b] Multidimensional Prognostic Index categories, age, gender (n=5 features).

[c] Multidimensional Prognostic Index categories, age, gender, BMI, anticholinergic risk score, laboratory data (n=15 features).

[d] Lab data=serum albumin, sodium, serum hemoglobin, C-reactive protein, creatinine, urea, urea-to-creatinine ratio, and estimated glomerular filtration rate.

[e] Multidimensional Prognostic Index domains, age, gender (n=10 features).

[f] Multidimensional Prognostic Index domains, age, gender, BMI, anticholinergic risk score, laboratory data (n=20 features).

[g] LR-MLE: logistic regression with maximum likelihood estimation.

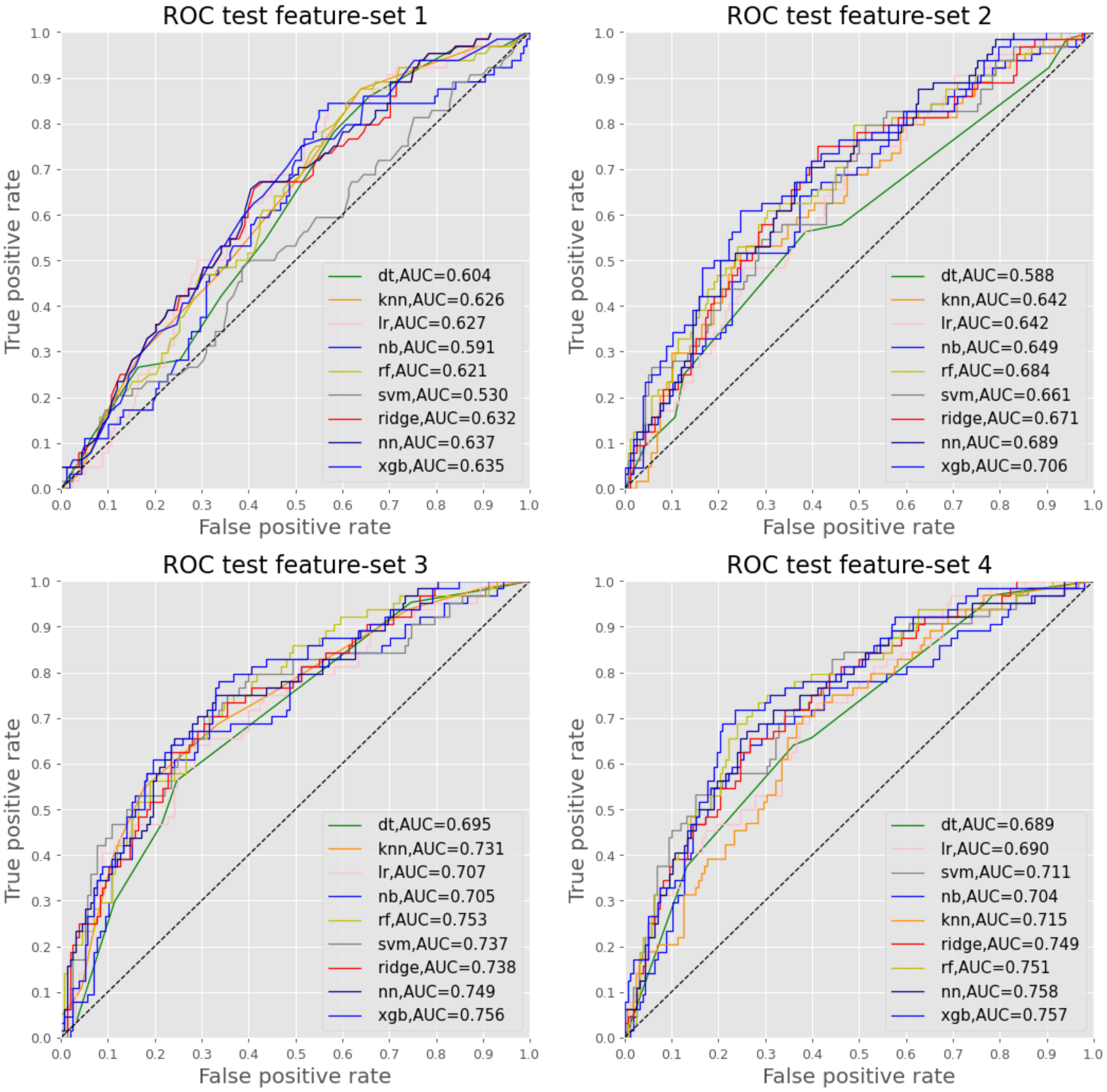
[h] XGB: extreme gradient boosting.

[i] Ridge: ridge regression.

[j] KNN: K-nearest neighbors.

[k] SVM: support vector machine.

Figure 2. Test accuracy of the 9 machine learning algorithms using feature sets 1 to 4. AUC: area under the receiver operating curve; dt: decision tree; knn: K-nearest neighbors; lr: logistic regression without penalization; nb: naive bayes; nn: neural network; rf: random forest; ridge: ridge regression; ROC: receiver operating curve; svm: support vector machine; xgb: eXtreme gradient boosting.



Training Data Accuracy

Multimedia Appendices 9 and 10 show the AUC values for LR-MLE and each ML algorithm for feature sets 1 to 4 using the training data set. The accuracy for LR-MLE was slightly higher for each feature set with the training data than for the test data values shown in Table 2. In comparison, the accuracy for each ML algorithm was considerably higher than that obtained using the test data, especially for the RF, XGB, and SVM, which obtained values of 0.956, 0.877, and 0.855, respectively, using feature set 4 and the training data.

Calibration

Multimedia Appendix 11 shows the distribution of predicted risk scores, the calibration plot, precision-recall curve, and receiver operating characteristic curve for the XGBoost algorithm using feature set 4. The calibration plot showed that overall, the predicted risks of mortality were in line with each observed risk decile. An AUC of 0.757 indicates fair to good accuracy in terms of overall sensitivity and specificity. The precision-recall curve indicates that the precision (ie, sensitivity or the ability to identify the patients that died) gradually decreased as the threshold for positivity decreased, and the recall, that is, the value of a positive classification, increased.

Feature Importance and Distributions

Figure 3 shows the feature importance plots for the XGBoost algorithm for the test data set using feature sets 1 to 4. In feature set 4, the MNA, IADL, and CIRS domains had the highest feature importance, indicating that they had the largest relative importance among the included features. Living alone (Cohab1) and urea were also highly ranked. The violin plots in Figure 4 show the distribution of MNA, IADL, and CIRS domains and urea—the four continuous features that accounted for the highest relative importance for the XGBoost algorithm in test feature set 4. The shape of the distributions was markedly different among patients who remained alive and those who died for the MNA score and the IADL score, which together accounted for 26.4% of the relative importance. The distributions were more similar for the CIRS score and urea, which accounted for 7.5% and 6.1% of the relative importance, respectively.

Figure 3. Feature importance plot for the eXtreme gradient boosting algorithm using test data with feature sets 1 to 4. ADL: activities of daily living; ARS: anticholinergic risk score; CIRS: Cumulative Illness Rating Scale; Cohab1: living alone; Cohab2: living with family or friends; Cohab3: living in an institute; Creat: creatinine; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; ESS: Exton Smith Scale; Hgb: serum hemoglobin; IADL: instrumental activities of daily living; MNA: Mini Nutritional Assessment; MPI: Multidimensional Prognostic Index; MNA: Mini Nutritional Assessment; ROC: receiver operating curve; SPMSQ: Short Portable Mental Status Questionnaire; Ur/Cr: urea-to-creatinine ratio.

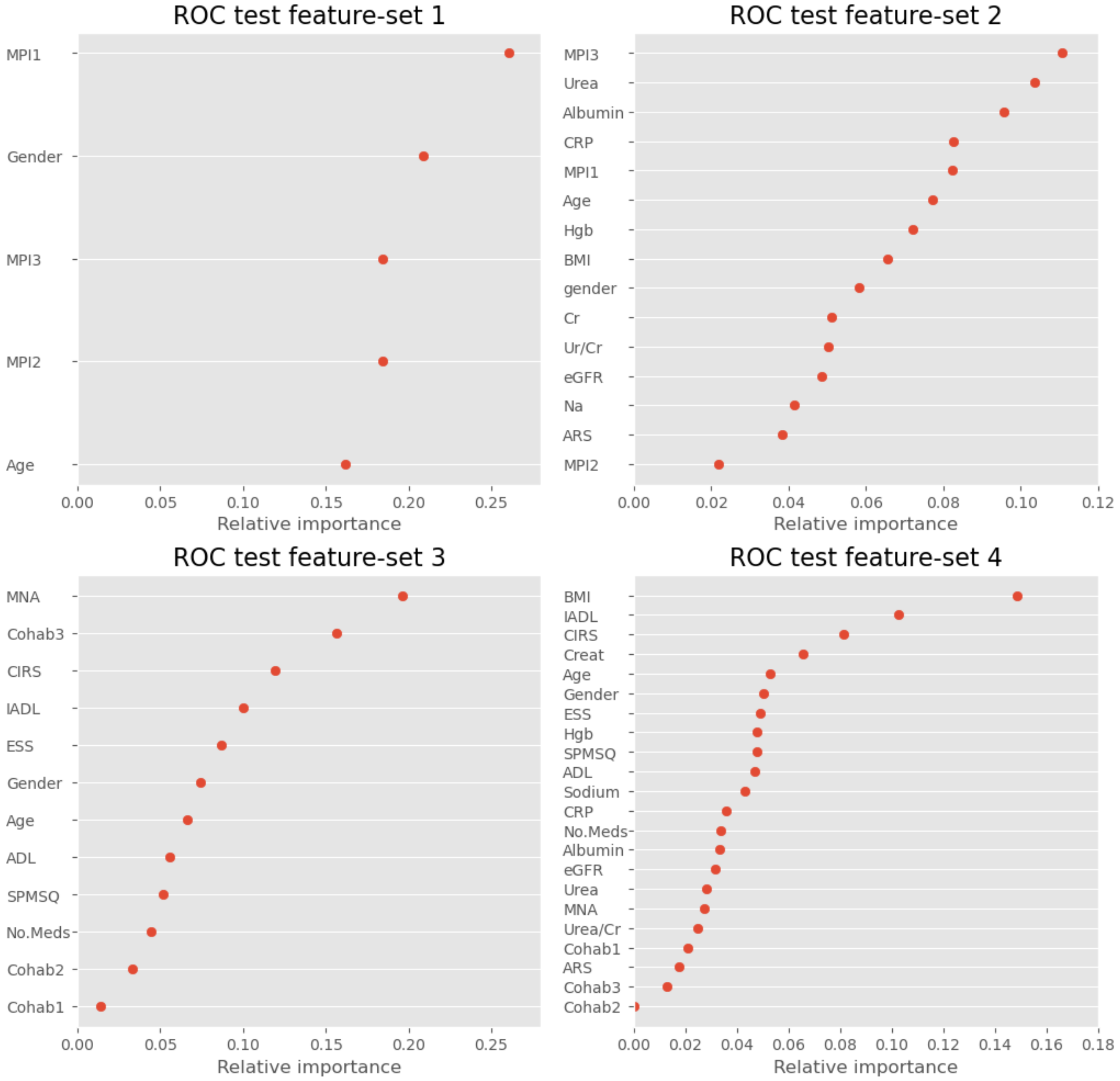
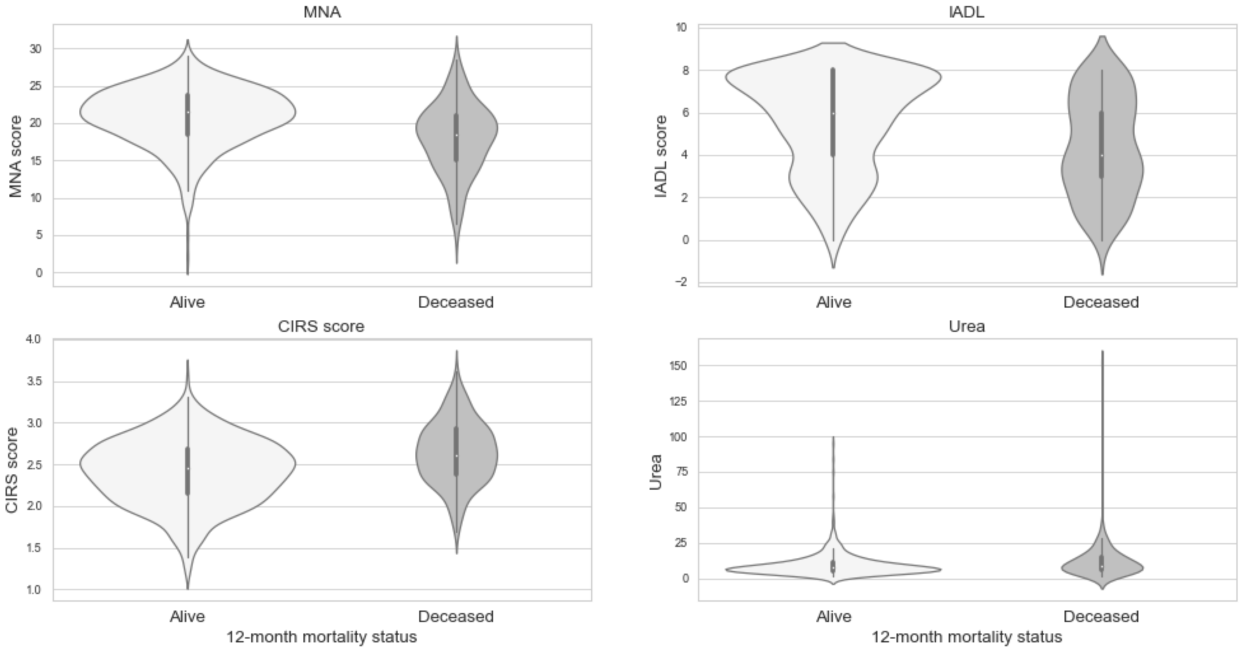


Figure 4. Violin plots showing distributions for the top 4 features for eXtreme gradient boosting in the second test feature set by patient vial status at 12 months after hospital discharge. CIRS: Cumulative Illness Rating Scale; IADL: instrumental activities of daily living; MNA: Mini Nutritional Assessment.



LR-MLE Accuracy

Table 3 shows the standardized odds ratios, 95% CIs, and *P* values for the LR-MLE model using test feature set 4. The variables that were statistically significant also had the largest standardized effect sizes: the MNA (OR 0.57, 95% CI 0.44-0.74; *P*<.001), CIRS domain (OR 1.81, 95% CI 1.32-2.49; *P*<.001), and the number of medications prescribed (OR 0.69, 95% CI 0.52-0.93; *P*=.02).

Table 3. Odds ratios (95% CIs) for the logistic regression with maximum likelihood estimation model using the test data with feature set 4[a].

Feature	Odds ratio (95% CI)	*P* value
Age	1.20 (0.90-1.61)	.21
ADL[b]	0.99 (0.71-1.39)	.96
IADL[c]	0.88 (0.63-1.22)	.44
SPMSQ[d]	0.99 (0.78-1.25)	.91
ESS[e]	0.89 (0.62-1.27)	.51
CIRS[f]	1.81 (1.32-2.49)	<.001
BMI	0.82 (0.63-1.07)	.14
MNA[g]	0.57 (0.44-0.74)	<.001
Sodium	0.96 (0.77-1.20)	.74
Urea	1.77 (0.89-3.51)	.10
Creatinine	0.85 (0.53-1.36)	.50
Albumin	0.83 (0.64-1.07)	.15
Hemoglobin	0.88 (0.68-1.14)	.35
Number of medications	0.69 (0.52-0.93)	.02
ARS[h]	1.06 (0.82-1.36)	.66
eGFR[i]	1.15 (0.74-1.80)	.53
CRP[j]	0.94 (0.73-1.20)	.61
Cohabitation		
Alone	1.00[k]	N/A[l]
Family or friends	0.96 (0.73-1.25)	.75
Institute	1.22 (0.95-1.57)	.12
Gender		
Female	1.00[k]	N/A
Male	0.69 (0.40-1.18)	.18
Urea-to-creatinine	1.13 (0.67-1.91)	.64

[a] All continuous variables were scaled before analysis to have a mean of zero and an SD of 1. Gender and cohabitation status were dummy coded for each category.

[b] ADL: activities of daily living.

[c] IADL: instrumental activities of daily living.

[d] SPMSQ: Short Portable Mental Status Questionnaire.

[e] ESS: Exton Smith Scale.

[f] CIRS: Cumulative Illness Rating Scale.

[g] MNA: Mini Nutritional Assessment.

[h] ARS: anticholinergic risk score.

[i] eGFR: estimated glomerular filtration rate.

[j] CRP: C-reactive protein.

[k] This is the reference group. Therefore, there is no CI.

[l] N/A: not applicable.

Sensitivity Analysis

In the sensitivity analysis that included all features as both first- and second-order terms, all results for the AUC for the nine ML algorithms were very similar to those obtained using only first order terms (data not shown).

Discussion

Principal Findings

In this study, we applied a range of ML binary-classification algorithms to data from 737 older inpatients in an Australian teaching hospital to develop and test a 12MM prediction model that can potentially be used in clinical settings to assist with risk management. The test data accuracy using the eight MPI domains and age and gender reached an AUC of 0.738 using LR-MLE, considerably higher than that obtained using the MPI categories plus age and gender (AUC=0.632). The addition of 10 clinical features improved the prediction accuracy further to AUC=0.757, which matched the accuracy obtained for the best overall ML-based algorithm (XGBoost), which outperformed most of the other algorithms except for the largest feature set in which the NN algorithm had similar accuracy (AUC=0.758).

The major strengths of our study include the use of both LR-MLE and a wide range of commonly used ML algorithms to compare the prediction accuracy for aggregate versus domain-based data. In the development of the ML algorithms, we also used a systematic ML framework, with a grid search of the hyperparameter space and 10-fold cross-validation for each algorithm. The much-improved accuracy warrants the calculation of individual patient risk scores, which, with appropriately developed technology platforms linked to the MPI domain and clinical information, can be used to better stratify patient risk and provide appropriate posthospital discharge surveillance and care [22,23].

Comparison With Prior Work

The prediction accuracy using the MPI categories and the test data set (AUC=0.632) was very similar to the poor accuracy obtained previously with follow-up on only 697 of the same patients (AUC=0.62), in which all records were used for assessing prediction accuracy rather than using separate training and testing data sets [5]. However, the higher accuracy obtained using the eight individual component features of the MPI combined with clinical data led to an accuracy similar to that originally reported for the three-category MPI within the original Italian MPI cohort [2]. In addition, the improvement in prediction accuracy with ML algorithms for some but not all feature sets provides general support for the use of ML in addition to LR-MLE when developing risk scores, at least for moderately sized data sets and feature numbers. The accuracy values obtained for precision, recall, and F1 score, which are less subject to variation than the AUC in imbalanced data sets, were of the same order for LR-MLE and the XGBoost algorithm and within the upper range of values obtained for these metrics.

The significantly higher accuracy obtained by using the separate domains of the MPI compared with using the 3-category aggregate MPI supports other studies in which the use of component domain data outperformed an aggregate score. In a meta-analysis of 6 studies comparing individual domain feature input to aggregate weighted scores for mortality and intensive care unit transfer, prediction using ML algorithms or multivariate regression with separate features considerably enhanced prediction compared with that obtained using the aggregate scoring systems (AUC=0.80 vs AUC=0.73) [6]. These findings suggest that caution should be used when employing aggregate risk scoring systems and the need to consider the underlying individual components. In addition, in our study, certain MPI domains, including the MNA, IADL, ESS, and CIRS, had either strong feature importance using the XGBoost algorithm or were strongly associated with 12MM using LR-MLE. Therefore, it may also be possible to obtain predictive accuracy similar to or better than that of the 3-category MPI feature sets in this study, by using data collected for only a subset of the eight MPI domains and by using the individual items for these specific domains. However, such an approach requires validation using data collected from additional retrospective and/or future prospective cohorts.

Unlike many ML algorithms that attempt to reduce the potential for overfitting and increased variance, the traditional LR-MLE approach to prediction modeling is not implicitly designed to deal with bias, multicollinearity, nonlinearity, or feature interactions. Thus, although the addition of features and model complexity generally improves a model's performance during training, a new and larger model does not guarantee similar improvements in model fit in the validation phase. Indeed, without some form of additional penalty term in the model's loss function to ensure that the training model is not being overfit, a decrease in testing accuracy is not uncommon [24]. When an additional *L2* penalty term was applied to the Logistic Regression classifier, the resulting ridge regression ML classification algorithm provided a slight increase in prediction accuracy for all four test feature sets (mean AUC of 0.698 for ridge regression vs AUC of 0.667 for nonregularized logistic regression). This reduction in strength (ie, regularization) of the estimated parameters for the ridge regression and the other ML algorithms during the training phase of model development may partly explain the higher test accuracy for LR-MLE in which there is no equivalent model training or an additional penalty term to help ensure that the LR-MLE coefficients are not inflated and that residual error is not overly reduced. Similarly, although many of the ML algorithms had better training performance than the LR-MLE, this did not translate into better test performance, suggesting that these ML algorithms often overfit the data during training, a problem that could potentially be solved with greater tuning of the algorithm's hyper-parameters.

A common criticism of ML approaches compared with standard statistical approaches is that they rely on blackbox algorithms in which the source of the improved performance is not readily transparent [25,26]. One method to overcome this lack of transparency is through the calculation of feature importance, which shows the relative importance of each feature in terms of reducing prediction error [26]. By using this method, we demonstrated the very high relative importance of the MNA and IADL domains of the MPI regarding risk for 12MM. Thus, although the ML algorithms do not generally provide a single specific effect estimate and CI for each variable or feature, our findings support other studies in which the MNA [27] and the IADL [28] have been shown to be independently and strongly associated with medium-term mortality in similar cohorts. These two features were also the strongest independent features in the LR-MLE model using feature set 4, emphasizing their importance in 12MM prediction.

The independent predictive ability of functional, cognitive, and psychological measures in predicting mortality in older hospitalized patients has been previously identified, with the IADL, mini mental state examination, and geriatric depression scale all shown to be independent predictors of mortality after risk adjustment for clinical characteristics [1]. However, the prediction accuracy in that study was modest (AUC=0.690), suggesting that complementary clinical data may have further improved prediction. The improvement in performance with the addition of blood biochemistry data, BMI, and ARS in our study highlights the independent predictive value of clinical data. Many of these features are known markers of malnutrition or frailty [29,30] and are associated with poorer outcomes in older patients [12,31,32]. Their independence with the MPI domains is demonstrated by weak univariate correlations, indicating their potential to add predictive power to that obtained from the domains. As each biochemical marker is routinely measured in hospital laboratories, it is highly feasible to incorporate them into hospital-based prediction algorithms.

Our results for the ML algorithms, based on a rigorous train-test validation approach with a grid search of hyper-parameters for each algorithm during training, demonstrate the potential of such methods in predicting mortality in clinical settings. In particular, the XGBoost algorithm, which is known as a superior and fast prediction ML algorithm, had the highest overall level of accuracy for the feature sets using the validation data set, and one of the highest levels of accuracy using the training data set, suggesting that the prediction accuracy observed for our cohort is likely to be repeatable in other cohorts with a similar patient profile. A distinguishing feature of the XGBoost algorithm is its approach of building a *strong learner* from an ensemble of *weak learners* (ie, separate DTs) by building successive trees, in which additional weight is placed on harder to predict subsets of the training data [33]. XGBoost uses a second-order Taylor series expansion to approximate the value of the loss function and incorporates regularization to avoid overfitting [21]. These features together ensure a fast solution as well as an unbiased estimate of prediction accuracy, both of which are important in building an automated risk prediction model at a system level within a clinical setting.

Limitations

Our study had several limitations. In using only routine geriatric assessment tools and basic biochemical, medication, and demographic data, it is likely that we were limited in our ability to obtain higher levels of predictive accuracy. The development of better prediction models should be possible using a richer set of features, such as additional information on demographics, comorbidities, laboratory test measurements, and medication type. Specifically, the Charlson Comorbidity Index, which determines the risk of medium-term mortality due to a specific set of 17 different comorbidities, is generally readily calculable using the secondary diagnosis codes captured in most administrative databases [34]. Similarly, routine hospital blood tests capture biomarker data associated with 12MM [35], and most hospitals capture medication data that can be used to assess the predictive capacity of medication type, dose, and polypharmacy. Other potentially predictive features include the number of hospitalizations and GP visits in the previous 12 months [36]. Our study data were also limited by its relatively small size, given that the test data set included only 222 patients and 64 deaths. ML methods have been found to generally require larger data sets, and particularly a higher number of events before stable measures of prediction performance are obtained in comparison with standard statistical models [37]. Despite this limitation, we still achieved moderate to good validation accuracy using data sets with a relatively limited number of events. In addition, although we were careful to use cross-validation throughout the model training process and to validate each model in a held-out test data set, we cannot be sure that our prediction accuracy will be the same in different cohorts of the same patient population or in different older populations such as persons living in aged care facilities. Further validation of the models in additional retrospective cohorts as well as prospective cohorts is required to ensure generalizability.

Conclusions

An MPI domain-based approach, together with clinical and demographic data, improved the prediction of mortality compared with a logistic regression model that used the aggregate MPI score. The ML algorithms in this study generally provided improved prediction accuracy compared with LR-MLE. These results build on previous work for the MPI and suggest that implementing risk scores based on MPI domains and clinical data with ML prediction models can be used to support clinical decision-making with respect to the medium-term follow-up care of older hospitalized patients.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The machine learning algorithms.
[DOCX File , 20 KB - jmir_v23i6e26139_app1.docx]

Multimedia Appendix 2
The dataset Multidimensional Prognostic Index feature set 4.
[XLSX File (Microsoft Excel File), 66 KB - jmir_v23i6e26139_app2.xlsx]

Multimedia Appendix 3
Python code for correlation matrix.
[PDF File (Adobe PDF File), 264 KB - jmir_v23i6e26139_app3.pdf]

Multimedia Appendix 4
Python code for XGBoost feature importance.
[PDF File (Adobe PDF File), 65 KB - jmir_v23i6e26139_app4.pdf]

Multimedia Appendix 5
Python code for receiver operating curve.
[PDF File (Adobe PDF File), 107 KB - jmir_v23i6e26139_app5.pdf]

Multimedia Appendix 6
Python code for violin plots.
[PDF File (Adobe PDF File), 281 KB - jmir_v23i6e26139_app6.pdf]

Multimedia Appendix 7
Python code for machine learning algorithms.
[PDF File (Adobe PDF File), 2187 KB - jmir_v23i6e26139_app7.pdf]

Multimedia Appendix 8
Diagnostic accuracy with the training data set.
[DOCX File , 21 KB - jmir_v23i6e26139_app8.docx]

Multimedia Appendix 9
Table S2. Accuracy, precision recall, F1 score.
[DOCX File , 21 KB - jmir_v23i6e26139_app9.docx]

Multimedia Appendix 10
Training accuracy of the 9 machine learning algorithms using feature sets 1 to 4.
[PNG File , 339 KB - jmir_v23i6e26139_app10.png]

Multimedia Appendix 11
Plots showing predicted risk distribution, calibration curve, precision-recall curve, and the area under the curve for the eXtreme gradient boosting algorithm using test data and feature set 4.
[PNG File , 127 KB - jmir_v23i6e26139_app11.png]

References

1. Inouye SK, Peduzzi PN, Robison JT, Hughes JS, Horwitz RI, Concato J. Importance of functional measures in predicting mortality among older hospitalized patients. J Am Med Assoc 1998 Apr 15;279(15):1187-1193. [doi: 10.1001/jama.279.15.1187] [Medline: 9555758]
2. Pilotto A, Ferrucci L, Franceschi M, D'Ambrosio LP, Scarcelli C, Cascavilla L, et al. Development and validation of a multidimensional prognostic index for one-year mortality from comprehensive geriatric assessment in hospitalized older patients. Rejuvenation Res 2008 Feb;11(1):151-161 [FREE Full text] [doi: 10.1089/rej.2007.0569] [Medline: 18173367]
3. Gregersen M, Hansen TK, Jørgensen BB, Damsgaard EM. Frailty is associated with hospital readmission in geriatric patients: a prognostic study. Eur Geriatr Med 2020 Oct;11(5):783-792. [doi: 10.1007/s41999-020-00335-w] [Medline: 32468431]
4. Volpato S, Bazzano S, Fontana A, Ferrucci L, Pilotto A, MPI-TriVeneto Study Group. Multidimensional prognostic index predicts mortality and length of stay during hospitalization in the older patients: a multicenter prospective study. J Gerontol A Biol Sci Med Sci 2015 Mar;70(3):325-331 [FREE Full text] [doi: 10.1093/gerona/glu167] [Medline: 25209253]
5. Bryant K, Sorich MJ, Woodman RJ, Mangoni AA. Validation and adaptation of the multidimensional prognostic index in an older Australian cohort. J Clin Med 2019 Nov 1;8(11):- [FREE Full text] [doi: 10.3390/jcm8111820] [Medline: 31683820]
6. Linnen DT, Escobar GJ, Hu X, Scruth E, Liu V, Stephens C. Statistical modeling and aggregate-weighted scoring systems in prediction of mortality and ICU transfer: a systematic review. J Hosp Med 2019 Mar;14(3):161-169 [FREE Full text] [doi: 10.12788/jhm.3151] [Medline: 30811322]
7. Bias VD. Encyclopedia of Machine Learning. New York, USA: Springer; 2010.

8. Amoedo A, Martnez-Costa MD, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: http://los40.com, http://los40principales.com, http://cadena100.es, http://europafm.es and http://kissfm.es. Radio J: Int Stud 2009 Feb 1;6(1):5-20 [FREE Full text] [doi: 10.1386/rajo.6.1.5_4]
9. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J 2017 Jun 14;38(23):1805-1814 [FREE Full text] [doi: 10.1093/eurheartj/ehw302] [Medline: 27436868]
10. Li R, Shinde A, Liu A, Glaser S, Lyou Y, Yuh B, et al. Machine learning-based interpretation and visualization of nonlinear interactions in prostate cancer survival. JCO Clin Cancer Inform 2020 Jul;4:637-646 [FREE Full text] [doi: 10.1200/CCI.20.00002] [Medline: 32673068]
11. Rudolph JL, Salow MJ, Angelini MC, McGlinchey RE. The anticholinergic risk scale and anticholinergic adverse effects in older persons. Arch Intern Med 2008 Mar 10;168(5):508-513. [doi: 10.1001/archinternmed.2007.106] [Medline: 18332297]
12. Cabrerizo S, Cuadras D, Gomez-Busto F, Artaza-Artabe I, Marín-Ciancas F, Malafarina V. Serum albumin and health in older people: review and meta analysis. Maturitas 2015 May;81(1):17-27. [doi: 10.1016/j.maturitas.2015.02.009] [Medline: 25782627]
13. Dong X, Mendes de Leon C, Artz A, Tang Y, Shah R, Evans D. A population-based study of hemoglobin, race, and mortality in elderly persons. J Gerontol A Biol Sci Med Sci 2008 Aug;63(8):873-878. [doi: 10.1093/gerona/63.8.873] [Medline: 18772477]
14. Chowdhury EK, Langham RG, Owen A, Krum H, Wing LM, Nelson MR, Second Australian National Blood Pressure Study Management Committeem. Comparison of predictive performance of renal function estimation equations for all-cause and cardiovascular mortality in an elderly hypertensive population. Am J Hypertens 2015 Mar;28(3):380-386. [doi: 10.1093/ajh/hpu160] [Medline: 25239479]
15. Holland-Bill L, Christiansen CF, Heide-Jørgensen Y, Ulrichsen SP, Ring T, Jørgensen JO, et al. Hyponatremia and mortality risk: a Danish cohort study of 279 508 acutely hospitalized patients. Eur J Endocrinol 2015 Jul;173(1):71-81. [doi: 10.1530/EJE-15-0111] [Medline: 26036812]
16. Alley DE, Crimmins E, Bandeen-Roche K, Guralnik J, Ferrucci L. Three-year change in inflammatory markers in elderly people and mortality: the Invecchiare in Chianti study. J Am Geriatr Soc 2007 Nov;55(11):1801-1807 [FREE Full text] [doi: 10.1111/j.1532-5415.2007.01390.x] [Medline: 17727645]
17. Lowry E, Woodman RJ, Soiza RL, Mangoni AA. Associations between the anticholinergic risk scale score and physical function: potential implications for adverse outcomes in older hospitalized patients. J Am Med Dir Assoc 2011 Oct;12(8):565-572. [doi: 10.1016/j.jamda.2011.03.006] [Medline: 21514242]
18. Jiang M, Zou Y, Xin Q, Cai Y, Wang Y, Qin X, et al. Dose-response relationship between body mass index and risks of all-cause mortality and disability among the elderly: A systematic review and meta-analysis. Clin Nutr 2019 Aug;38(4):1511-1523. [doi: 10.1016/j.clnu.2018.07.021] [Medline: 30082166]
19. Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. JAMA Netw Open 2020 Jan 3;3(1):e1919396 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.19396] [Medline: 31951272]
20. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. GetMobile: Mobile Comp 2015 Jun;19(1):29-33. [doi: 10.1145/2786984.2786995]
21. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Conference on Knowledge Discovery and Data. 2016 Presented at: KDD'16; June 8-11, 2016; Alaska, USA.
22. Brajer N, Cozzi B, Gao M, Nichols M, Revoir M, Balu S, et al. Prospective and external evaluation of a machine learning model to predict in-hospital mortality of adults at time of admission. JAMA Netw Open 2020 Feb 5;3(2):e1920733 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20733] [Medline: 32031645]
23. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. Lancet 2021 Jan;397(10270):199-207. [doi: 10.1016/s0140-6736(20)32519-8]
24. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York, USA: Springer; 2009.
25. Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning ‐ based prediction models in healthcare. WIREs Data Mining Knowl Discov 2020 Jun 29;10(5):-. [doi: 10.1002/widm.1379]
26. Elshawi R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. BMC Med Inform Decis Mak 2019 Jul 29;19(1):146 [FREE Full text] [doi: 10.1186/s12911-019-0874-0] [Medline: 31357998]
27. Lundin H, Sääf M, Strender L, Mollasaraie HA, Salminen H. Mini nutritional assessment and 10-year mortality in free-living elderly women: a prospective cohort study with 10-year follow-up. Eur J Clin Nutr 2012 Sep;66(9):1050-1053. [doi: 10.1038/ejcn.2012.100] [Medline: 22947901]
28. Hennessy S, Kurichi JE, Pan Q, Streim JE, Bogner HR, Xie D, et al. Disability stage is an independent risk factor for mortality in medicare beneficiaries aged 65 years and older. PM R 2015 Dec;7(12):1215-1225 [FREE Full text] [doi: 10.1016/j.pmrj.2015.05.014] [Medline: 26003869]

29. Saedi AA, Feehan J, Phu S, Duque G. Current and emerging biomarkers of frailty in the elderly. Clin Interv Aging 2019;14:389-398 [FREE Full text] [doi: 10.2147/CIA.S168687] [Medline: 30863033]
30. Mailliez A, Guilbaud A, Puisieux F, Dauchet L, Boulanger. Circulating biomarkers characterizing physical frailty: CRP, hemoglobin, albumin, 25OHD and free testosterone as best biomarkers. Results of a meta-analysis. Exp Gerontol 2020 Oct 1;139:111014. [doi: 10.1016/j.exger.2020.111014] [Medline: 32599147]
31. Denny SD, Kuchibhatla MN, Cohen HJ. Impact of anemia on mortality, cognition, and function in community-dwelling elderly. Am J Med 2006 Apr;119(4):327-334. [doi: 10.1016/j.amjmed.2005.08.027] [Medline: 16564775]
32. O'Hare AM, Choi AI, Bertenthal D, Bacchetti P, Garg AX, Kaufman JS, et al. Age affects outcomes in chronic kidney disease. J Am Soc Nephrol 2007 Oct;18(10):2758-2765 [FREE Full text] [doi: 10.1681/ASN.2007040422] [Medline: 17855638]
33. Friedman J, Popescu B. Importance Sampled Learning Ensembles. Statistics. 2003. URL: https://statweb.stanford.edu/~jhf/ftp/isle.pdf [accessed 2021-06-01]
34. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992 Jun;45(6):613-619. [doi: 10.1016/0895-4356(92)90133-8] [Medline: 1607900]
35. Kristensen M, Iversen AK, Gerds TA, Østervig R, Linnet JD, Barfod C, et al. Routine blood tests are associated with short term mortality and can improve emergency department triage: a cohort study of >12,000 patients. Scand J Trauma Resusc Emerg Med 2017 Nov 28;25(1):115 [FREE Full text] [doi: 10.1186/s13049-017-0458-x] [Medline: 29179764]
36. Gray DJ, Sidaway-Lee K, White E, Thorne A, Evans PH. Continuity of care with doctors-a matter of life and death? A systematic review of continuity of care and mortality. BMJ Open 2018 Jun 28;8(6):e021161 [FREE Full text] [doi: 10.1136/bmjopen-2017-021161] [Medline: 29959146]
37. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014 Dec 22;14:137 [FREE Full text] [doi: 10.1186/1471-2288-14-137] [Medline: 25532820]

Abbreviations

12MM: 12-month mortality
ADL: activities of daily living
ARS: anticholinergic risk score
AUC: area under the receiver operating curve
CGA: comprehensive geriatric assessment
CIRS: Cumulative Illness Rating Scale
CRP: C-reactive protein
DT: decision tree
eGFR: estimated glomerular filtration rate
ESS: Exton Smith Scale
IADL: instrumental activities of daily living
KNN: K-nearest neighbors
LR-MLE: logistic regression with maximum likelihood estimation
ML: machine learning
MNA: Mini Nutritional Assessment
MPI: Multidimensional Prognostic Index
NN: neural network
OR: odds ratio
RF: random forest
SVM: support vector machine
XGBoost: eXtreme gradient boosting

Edited by R Kukafka; submitted 01.12.20; peer-reviewed by E Damsgaard, E Scruth, JA Benítez-Andrades; comments to author 05.01.21; revised version received 16.01.21; accepted 13.05.21; published 21.06.21.

Please cite as:
Woodman RJ, Bryant K, Sorich MJ, Pilotto A, Mangoni AA
Use of Multiprognostic Index Domain Scores, Clinical Data, and Machine Learning to Improve 12-Month Mortality Risk Prediction in Older Hospitalized Patients: Prospective Cohort Study
J Med Internet Res 2021;23(6):e26139
URL: https://www.jmir.org/2021/6/e26139
doi:10.2196/26139
PMID:34152274

Original Paper

Design and Implementation of a Real-time Monitoring Platform for Optimal Sepsis Care in an Emergency Department: Observational Cohort Study

Andy Hung-Yi Lee[1], MD, MBA; Emily Aaronson[1], MD, MPH; Kathryn A Hibbert[2], MD; Micah H Flynn[3], BA, MM; Hayley Rutkey[3], BSc; Elizabeth Mort[2], MD, MPH; Jonathan D Sonis[1], MD; Kyan C Safavi[3], MD, MBA

[1]Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, United States

[2]Department of Medicine, Massachusetts General Hospital, Boston, MA, United States

[3]Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Andy Hung-Yi Lee, MD, MBA
Department of Emergency Medicine
Massachusetts General Hospital
55 Fruit Street
Boston, MA, 02114
United States
Phone: 1 617 724 4100
Email: alee85@partners.org

Abstract

Background: Sepsis is the leading cause of death in US hospitals. Compliance with bundled care, specifically serial lactates, blood cultures, and antibiotics, improves outcomes but is often delayed or missed altogether in a busy practice environment.

Objective: This study aims to design, implement, and validate a novel monitoring and alerting platform that provides real-time feedback to frontline emergency department (ED) providers regarding adherence to bundled care.

Methods: This single-center, prospective, observational study was conducted in three phases: the design and technical development phase to build an initial version of the platform; the pilot phase to test and refine the platform in the clinical setting; and the postpilot rollout phase to fully implement the study intervention.

Results: During the design and technical development, study team members and stakeholders identified the criteria for patient inclusion, selected bundle measures from the Center for Medicare and Medicaid Sepsis Core Measure for alerting, and defined alert thresholds, message content, delivery mechanisms, and recipients. Additional refinements were made based on 70 provider survey results during the pilot phase, including removing alerts for vasopressor initiation and modifying text in the pages to facilitate patient identification. During the 48 days of the postpilot rollout phase, 15,770 ED encounters were tracked and 711 patient encounters were included in the active monitoring cohort. In total, 634 pages were sent at a rate of 0.98 per attending physician shift. Overall, 38.3% (272/711) patients had at least one page. The missing bundle elements that triggered alerts included: antibiotics 41.6% (136/327), repeat lactate 32.4% (106/327), blood cultures 20.8% (68/327), and initial lactate 5.2% (17/327). Of the missing Sepsis Core Measures elements for which a page was sent, 38.2% (125/327) were successfully completed on time.

Conclusions: A real-time sepsis care monitoring and alerting platform was created for the ED environment. The high proportion of patients with at least one alert suggested the significant potential for such a platform to improve care, whereas the overall number of alerts per clinician suggested a low risk of alarm fatigue. The study intervention warrants a more rigorous evaluation to ensure that the added alerts lead to better outcomes for patients with sepsis.

(J Med Internet Res 2021;23(6):e26946) doi:10.2196/26946

KEYWORDS

electronic monitoring platform; sepsis; quality improvement

Introduction

Sepsis Background

Sepsis is the leading cause of death in US hospitals [1], accounting for 6% of all hospitalizations and 35% of all inpatient deaths [2]. Hospital performance in sepsis care has a significant impact on patient outcomes. A recent study demonstrated that every hour of delay in the completion of a sepsis bundle, including antibiotic administration, was associated with a 4% increase in risk-adjusted hospital mortality [3]. Thus, every failure to complete a sepsis bundle on time potentially represents a barrier to optimal patient care. There is a clear and urgent need for effective interventions to improve the performance of sepsis care and meet the expectations of patients and stakeholders.

International guidelines and the Centers for Medicare and Medicaid Services (CMS) Sepsis Core Measures (SEP-1) bundle emphasize the importance of adhering to specific steps in the diagnosis and management of sepsis [4]. Consistent adherence can be very challenging, especially in the setting of a busy emergency department (ED), ward, or intensive care unit, where there are multiple simultaneous demands on providers' attention. In this environment, SEP-1 bundle care can be delayed or missed, even when team members are knowledgeable about best practices in sepsis care [5].

The CMS SEP-1 severe sepsis bundle is an all-or-nothing measure that requires antibiotics, blood cultures, and a lactate measurement within 3 hours of sepsis onset and a repeat lactate measurement within 6 hours if the initial lactate level is elevated. In cases of septic shock, additional requirements include adequate fluid resuscitation, reassessment of volume status and perfusion, and possible addition of vasopressors. As an all-or-nothing measure, a deficiency in any element of the bundle counts as failing the entire measure. In the first 2 years after the measure was introduced in 2015, two-thirds of sepsis cases submitted to CMS by hospitals failed the measure [6], and recent Hospital Compare data from 2019 still show a failure rate of 61% [7]. There has been argument in the literature about the appropriateness of some components of the bundle [8-10]; however, the fact remains that early identification and treatment of patients with sepsis remains an important quality opportunity.

Strategies to Improve Sepsis Care

Institutions have attempted both nondigital and digital processes to improve sepsis care, including provider and nursing huddles and standardized electronic order sets. Most digital initiatives studied thus far have focused on identifying patients at risk for sepsis using the severity of illness scores in the electronic medical record (EMR). This has helped improve the early detection of sepsis [11,12] but has only modest effects on adherence to evidence-based bundle care elements and outcomes [13]. In concept, digital solutions offer the potential to close this gap by enabling real-time monitoring of adherence and providing just-in-time alerts to providers at the bedside. Electronically available data from the patient's medical record would make this possible, as it includes most of the key elements of the SEP-1 bundle, including venipuncture for laboratory studies and medication administration. Any solution that could increase adherence and decrease cognitive load must also ensure that it does not cause alarm fatigue.

To date, we have not found any studies in the literature describing electronic alerting based on the completion of bundle elements for sepsis. Therefore, we developed a digital software solution called the Sepsis Care Tracking Platform (SCTP), which continuously monitors patients at risk of sepsis for completion of SEP-1 bundle elements and delivers actionable reminders to medical providers at the bedside in time for them to correct these deficiencies.

Study Objectives

The objective of this paper is to describe the design, implementation, and validation of a novel electronic monitoring platform for optimal sepsis care in an ED at a large urban teaching hospital. By establishing feasibility, we hope to lay the groundwork for future randomized controlled studies that can determine the efficacy of the electronic monitoring platform in impacting process and outcome measures in sepsis care.

Methods

Initial Stakeholder Consultation

Multiple stakeholder groups were consulted to devise and implement the SCTP in the ED. For project approval and scope, the study team met with the chairperson and operational leaders of the Department of Emergency Medicine. The project was also presented to the quality and safety committees, both at the departmental and hospital levels. For feedback and iterative improvement, the ED Quality and Safety Committee was updated every 2 weeks regarding the project status and areas for improvement. ED providers who were paged by the SCTP received surveys regarding their experience and suggestions for improvement. This study was a quality improvement project based on institutional guidelines from the institutional review board and was thus exempted from the institutional review board review.

Project Overview

The timeline for this project was divided into three phases: design and technical development, pilot, and postpilot program rollout.

Design and Technical Development Phase

The design and technical development phase spanned from September 1 to November 30, 2019. During this phase, the study team aimed to develop components of the SCTP and alert workflow, including defining the patient cohort, selecting the bundle measures from CMS SEP-1 for alerting, and determining recipients of alerts, mode of alert delivery, and alert content. In addition, technical development was performed, including linking EMR data to the platform, validating that the correct patients and bundle elements were captured, and performing initial quality control on data and alerts via chart review.

Pilot Phase

The pilot phase spanned from December 9, 2019, to January 12, 2020. During this phase, the study team deployed the initial version of the platform in the ED, including monitoring actual patients and sending alerts to their providers. This phase aimed to ensure the technical stability of the platform through confirmed receipt of alerts, monitor for false-positive results, and obtain feedback from alert recipients via a survey to determine the usefulness of the platform, alarm fatigue, and potential improvements. Monitoring for false positives was performed by study team members who manually verified every page sent during the pilot phase by confirming through the patient record that alerts reflected the correct bundle element deficiencies. Providers who received a page would be emailed a two-question survey the next day asking for feedback regarding each page. The two questions in the survey were (1) *Was this page alert useful in providing sepsis care to the patient?* with a ranking from *not at all helpful* (1) to *extremely helpful* (5), and (2) *If not helpful, why?* An additional space allowed for additional feedback, and providers could receive the survey multiple times for different pages.

Postpilot Rollout Phase

After incorporating modifications and improvements identified during the pilot phase, the postpilot rollout was initiated, in which a prospective trial was performed between January 13, 2020, and March 1, 2020. This trial aimed to assess the function of the platform and describe alerting patterns and subsequent adherence to alerts during this period.

Platform Setup

The SCTP ED platform was coded in C#, using .NET Microsoft Development Platform and Microsoft SQL databases. It was linked to the institutional EMR (EPIC) through a web service that extracted data each minute. The SCTP checked for the completion of the sepsis bundle elements every 15 minutes. These elements included completion of blood culture orders; initial and repeat lactate results; and administration of antibiotics, vasopressors, and fluids (as described in the *Results* section, vasopressors and fluids were later eliminated from the SCTP). If one or more elements were incomplete, the ED providers would be paged with the deficiency. Each patient was added to the platform after the institutional sepsis best practice alert (BPA) was accepted and monitored until the end of the 6-hour window for optimal sepsis care. The platform tracked a patient within that window even if the patient was admitted or transferred to a different area of the hospital and continued to page the appropriate responding clinician.

Study Measures and Analysis

During both the pilot and postpilot program rollout phases, several study measures were tracked to assess the platform function and outcomes. Data were abstracted from the SCTP, an institutional clinical and administrative database, and surveys administered to paged providers. Subsequent data analyses were performed using Microsoft Excel.

We defined an *alert* as when the SCTP identified a sepsis bundle element deficiency within 1 hour of the bundle element becoming due. We defined a *page* as occurring when a clinical provider received a notification on their pager. As there were usually multiple providers caring for each patient, a single *alert* could trigger multiple *pages* at the same time. During the pilot phase, we assessed the total number of pages sent to providers; the number of alerts per provider shift, defined as the number of alerts divided by the number of shifts completed by the attending providers; and the etiology of alerts, defined as the bundle measure that triggered the alert. In addition, we administered an email survey to the providers after receiving an alert. The survey consisted of two questions regarding whether the pages were helpful and whether they had any further feedback regarding the SCTP.

During the postpilot program rollout, we measured the characteristics of patients who were monitored by the platform, including age and length of stay in the ED. Similar to the pilot phase, we also measured the number and etiology of the pages sent to the providers. In addition, we measured the *postalert successful adherence rate* of sepsis bundle measures, which was defined as the completion of the sepsis bundle element within the next hour after the relevant page was sent to the provider.

Results

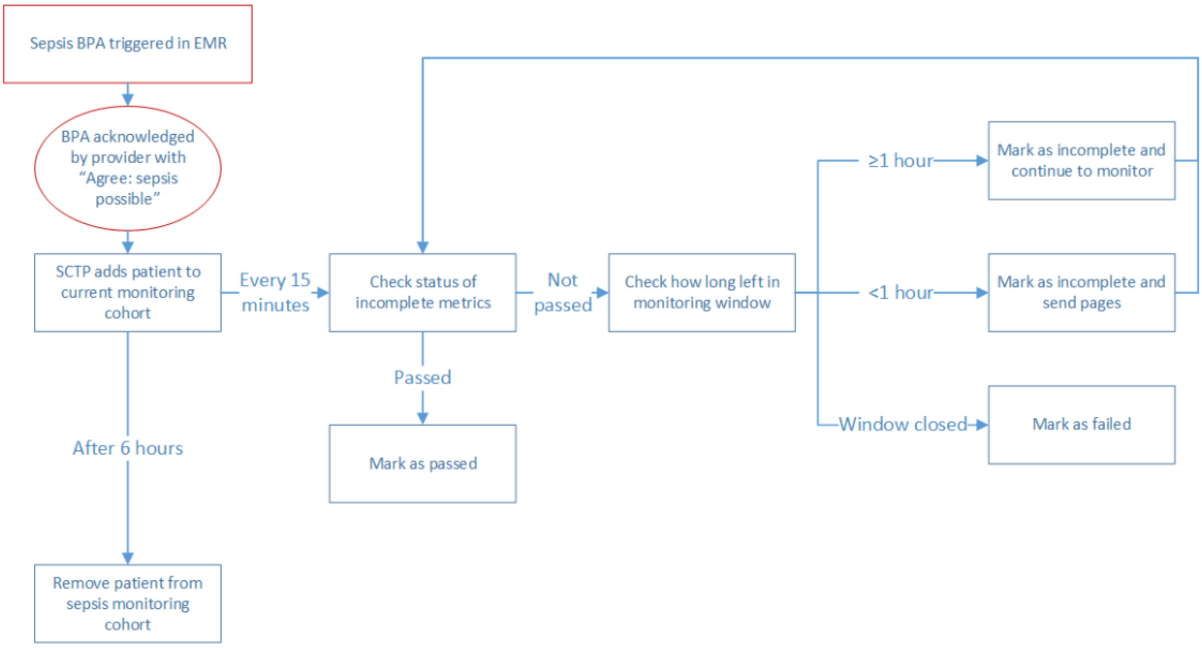
SCTP Design and Technical Development Phase

The SCTP was developed and validated in close cooperation with ED Quality and Safety leadership and intended to fit within the existing workflow of the institution to decrease barriers to adoption.

Inclusion Criteria for Active Monitoring

In consultation with ED Quality and Safety leadership, it was determined that patients should qualify for monitoring if they met the criteria for possible sepsis based on an electronic health record (EHR)–derived algorithm that existed before this study. This algorithm is based on the Sepsis-3 2016 European Society of Intensive Care Medicine and Society of Critical Care Medicine consensus definition of sepsis and incorporates both signs of infection such as fever, positive laboratory results, and patient risk factors, as well as signs of organ dysfunction as measured by the Sequential Organ Failure Assessment score [14]. Providers caring for patients identified as potentially septic receive a prompt in the EHR in which they could click that they agreed or disagreed that sepsis was likely. If accepted, the patient would qualify for SCTP monitoring. Stakeholders and study team members believed that the advantage of this approach was to include only those patients in the active monitoring cohort whom the team suspected of sepsis, thereby increasing the relevance of SCTP alerts they received (Figure 1).

Figure 1. Sepsis care tracking platform page generation logic. BPA: best practice alert; EMR: electronic medical record; SCTP: Sepsis Care Tracking Platform; SEP-1: sepsis core measure.



Measurement Selection for Alerting

From the commonly accepted SEP-1 bundle, stakeholders selected antibiotics, blood cultures, and lactate measurements as the bundle elements that would be monitored and excluded measures related to the volume of intravenous fluids administered. The decision to exclude fluids was made because of the belief that significant practice variation existed due to patient-specific conditions and comorbidities [15,16] and that alerts suggesting that additional fluids should be administered could be clinically inappropriate.

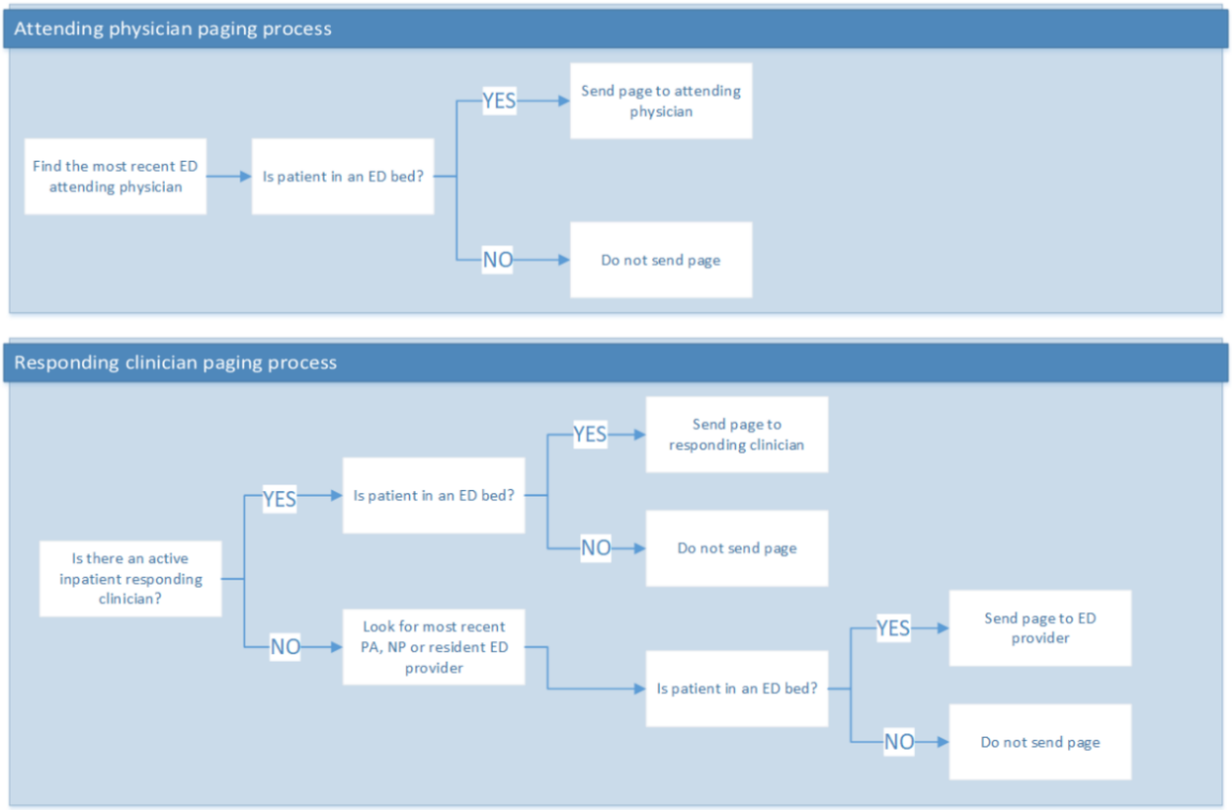
Alert Time Threshold

Stakeholders agreed that paging 1 hour before the deadline for bundle compliance gave providers sufficient time to address the deficiency (Figure 1). To minimize alert fatigue, stakeholders agreed that providers would receive a maximum of two pages per patient that would be sent 1 hour before the 3-hour and 6-hour SEP-1 time windows. Each page would list all the deficient bundle elements at that time.

Alert Recipients

Given that multiple providers were simultaneously responsible for a single patient, stakeholders designed the software to use the logic shown in Figure 2 to determine the appropriate clinicians to page. The decision was made to send alerts to both the responding clinician (resident or advanced practitioner) and the supervising attending because each played a unique role in the patient's sepsis care. The resident and advanced practitioner were responsible for placing orders and closing the loop of communication with bedside nurses about bundle measures, whereas the attending physician could supervise and monitor the overall care of the patient to ensure appropriate care was provided. Although the study team and ED stakeholders thought it valuable to alert bedside nurses, given their important role in sepsis care delivery, this was technically not feasible as they did not carry pager devices.

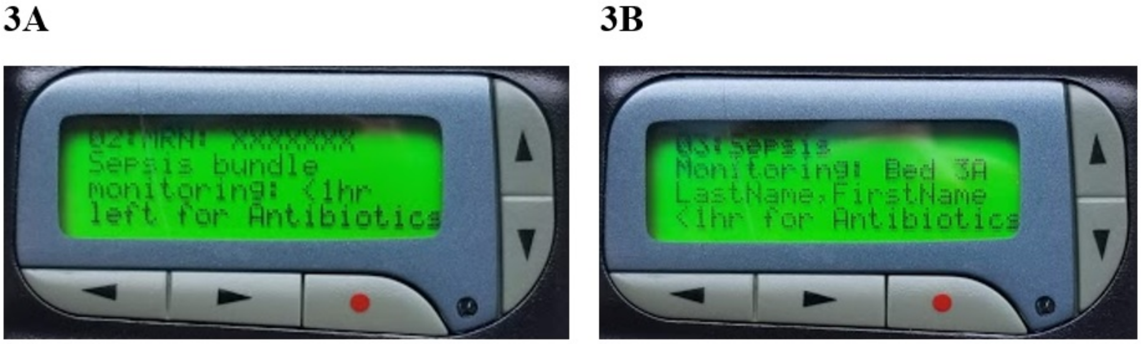
Figure 2. Sepsis care tracking platform page recipient logic. ED: emergency department; NP: nurse practitioner; PA: physician assistant.



Alert Message Contents

Stakeholders agreed that alert message contents should state the deficient element without providing prescriptive language recommending that the provider performs a particular action (Figure 3). The message contents would list which element was deficient, but otherwise the remainder of the message would remain the same for all pages. This decision was made on the rationale that in some cases providers would appropriately not administer bundle care to patients based on the evolving diagnostic and clinical course of the patient.

Figure 3. Pilot phase (3A) and rollout phase (3B) pager display.



Alert Delivery Mechanism

Stakeholders determined that the hospital's paging system was the most effective means of communicating time-sensitive alerts, given that providers always carry their pager and typically view messages immediately. This was preferred over EMR-based pop-up alerts because it required the provider to log on to the patient's medical record to see the alerts. It was felt that this mode would be burdensome to the provider and potentially delay the receipt of the message.

Pilot Phase

During the 35-day period, 453 patients met the inclusion criteria to be monitored by the SCTP. Overall, 371 pages were sent, resulting in 0.79 pages per attending shift. Of the alerts, 27.3% (50/183) were for blood cultures, 50.3% (92/183) were for antibiotics, and 22.4% (41/183) were for lactate measurements.

In total, 70 surveys were returned with responses. In total, 87% (61/70) of respondents found the alerts to be helpful (21/70, 30%), very helpful (17/70, 24%), or extremely helpful (23/70, 33%). Specific examples of positive feedback included "The page helped us realize blood cultures were not initially ordered and we ordered them after this page" and "It prompted us to reassess the patient for sepsis." Providers also suggested that pages "would be more useful if patient's name was also displayed on the page and not only the MRN." Providers also noted that their suspicion for sepsis can change over the course of an ED visit, with comments such as "Patient was later felt not to have sepsis so bundle was not completed." On the basis of this feedback, the alert message content was modified to allow providers to more easily identify the patient (Figure 3) and the alert routing algorithm was adjusted to identify the correct attending physician (Figure 2). We also removed reminders to initiate vasopressors for hypotensive patients because a reliable bedside monitoring alarm system already existed to alert clinicians to hypotension and clinicians consistently responded by initiating vasopressors when appropriate. Our survey responses helped to elucidate this frontline practice.

Postpilot Program Rollout Phase

After completion of the pilot program with subsequent adjustments to the SCTP, the postpilot program rollout phase was initiated. During this 48-day period, 711 patient encounters (mean 14.8, SD 4.4 encounters per day) for 648 unique patients met the inclusion criteria to be monitored by the SCTP (Table 1).

Of the 711 encounters monitored by the SCTP, 439 (61.7%) passed all measures without requiring a page, as all sepsis bundle elements were completed more than an hour before the deadline. In comparison, 38.3% (272/711) patients had one or more pages sent to the responding providers. In total, 634 pages were sent, averaging 0.89 pages per patient in the active monitoring cohort and 0.98 pages per attending shift (Table 2).

Of the 634 pages, the most frequent alert sent was for deficient antibiotics (136/327, 41.6%), followed by the second lactate measurement (106/327, 32.4%), blood cultures (68/327, 20.8%), and first lactate (17/327, 5.2%). This distribution was similar to the results from the pilot study, with pages related to antibiotics being the most common etiology.

Overall, there was a substantial proportion of sepsis bundle element deficiencies that triggered a page and were successfully adhered to within the SEP-1 time window. There was a 26% (18/68) postalert successful adherence rate for collecting blood cultures, 12.5% (17/136) for administering antibiotics, and 24% (4/17) for initial lactate measurement and an impressive 81.1% (86/106) for the second lactate measurement.

Table 1. Postpilot rollout phase patient demographics (N=648).

Characteristics	Values, n (%)
Age (years)	
18-65	271 (41.8)
>65	377 (58.2)
Sex	
Female	294 (45.4)
Male	354 (54.6)
ICU[a] admission	
No	468 (72.2)
Yes	180 (27.8)
Emergency department LOS[b] (hours)	
≤6	236 (36.4)
>6	339 (52.3)

[a] ICU: intensive care unit.

[b] LOS: length of stay.

Table 2. Postpilot rollout phase paging and alert data.

Characteristics	Values
Emergency department encounters	15,770
Encounters in active monitoring (per day), mean (SD)	711 (14.8)
Total pages (per day), mean (SD)	634 (13.2)
Pages per attending shift	0.98
Encounters with at least one page (n=711), n (%)	272 (38.3)
Encounters with no pages (n=711), n (%)	439 (61.7)
Alert etiology (n=327), n (%)	
Total	327 (100)
Blood cultures	68 (20.8)
Antibiotics	136 (41.6)
Initial lactate	17 (5.2)
Repeat lactate	106 (32.4)
Postalert successful adherence (n=327), n (%)	
Total	125 (38.2)
Blood cultures	18 (26.5)
Antibiotics	17 (12.5)
Initial lactate	4 (23.5)
Repeat lactate	86 (81.1)

Discussion

Principal Findings

To improve adherence to sepsis guideline care at our tertiary care hospital, we designed and implemented a real-time sepsis care tracking and alerting platform in a busy ED environment. With feedback from frontline clinicians, quality and safety program representatives, and departmental leadership, we customized the SCTP to the needs of clinicians and the clinical environment. In our postpilot program rollout phase, the SCTP actively monitored 711 ED encounters where sepsis was thought to be likely, in 272 of these encounters, the platform sent text pages to ED providers due to a deficiency in sepsis bundle completion, with an overall postpage adherence rate of 38.2% (125/327).

During the development phase, it was important to select the sepsis bundle elements that were felt to be most clinically appropriate and could be identified by our monitoring platform. The SEP-1 framework served as a foundation for this study, as the framework was widely accepted and aligned the interests of multiple stakeholder groups, including providers, departmental and quality and safety leadership, and hospital administration. Engaging with provider stakeholders allowed us to make the application more effective. Through these discussions, we were able to understand the clinical workflow and target alerting to the frontline provider who places the orders.

Moreover, we note the importance of the provider feedback obtained during the pilot phase. The surveys emphasized the need to provide critical information concisely in the SCTP paging alerts, including what information can quickly and accurately identify the patient and the specific bundle element deficiency. In addition, given the complex structure of provider teams in our ED, which can change multiple times over a patient's ED visit, we had to consider different clinical scenarios and which provider would be the optimal recipient for paging alerts. Finally, given the learning curve that comes with new technologies, the pilot phase allowed providers to acclimate to this system before the full rollout.

During the postpilot rollout, we found that there was a substantial proportion of encounters (272/711, 38.3%) in which the clinician was at risk of missing the SEP-1 care measure and for which an alert was paged, indicating the potential impact of such a platform on adherence. Furthermore, we found that the postalert successful adherence rate ranged from 12.5% (17/136) to 81.1% (86/106), depending on the measure for which the alert was sent, further suggesting the potential for the platform to improve adherence. Finally, we found that the number of pages per clinician shift (0.98) was low enough to not be overly disruptive but high enough that we would expect the SCTP to meaningfully impact SEP-1 performance. On the basis of these findings, we believe that a real-time sepsis care tracking and alerting platform is feasible in a hectic ED environment and could have a significant impact on adherence to life-saving sepsis care without causing alarm fatigue.

In addition, we found it concerning that the most common deficiency triggering a paging alert was a delay in antibiotics, which is the SEP-1 element that is most closely linked to mortality in previous studies [17]. This suggests the possibility that the SCTP could be associated with mortality improvement if it is able to improve timely antibiotic administration. However, we found that the postalert adherence success rate for antibiotics was the lowest among all alerts. It may be that antibiotics were, in fact, administered just outside of the monitored 3-hour window or that patients were subsequently determined to likely not have sepsis. Regardless, this discrepancy requires further exploration to determine why antibiotics may have a lower rate of postalert adherence compared with other measures.

In comparison, we found that repeat lactate was the second most common etiology of pages (106/327, 32.4%) and had the highest postadherence rate (86/106, 81.1%). This suggests that the SCTP may be especially valuable over a prolonged time course in which providers may be distracted by the needs of other patients, and paging alerts in this situation can refocus provider attention with sufficient time to deliver optimal care.

We believe that the postalert adherence rate of 12.5% (17/136) to 81.1% (86/106) for bundle element deficiencies represents the completion of bundle elements that likely would have been missed otherwise and is a significant potential improvement in increasing adherence to sepsis bundle care. However, the majority of pages did not result in adherence, which deserves further study. On the basis of our survey data, we believe that an important driver of this trend may be that the clinician no longer believed sepsis was likely. The sepsis BPA, which the SCTP used as a trigger to begin monitoring, was designed to cast a wide net early in the patient's ED encounter to reduce the likelihood of missing any patient who could be septic given the high mortality of the disease. As more data are gathered and the patient's care evolves, clinical teams may have judged that sepsis was unlikely and thus appropriately withheld sepsis bundle care. We believe that this could have accounted for much of the nonadherence. A future iteration of the SCTP may enable frontline clinicians to *off ramp* from the page alerts if they judge that sepsis is highly unlikely later in the patient's ED course.

SCTP differs from previous EHR-related innovations in improving sepsis care. The majority of previous studies have focused on identifying sepsis through different clinical criteria [18] or through predictive algorithms and machine learning [19], usually followed by an EHR warning or paging alert. Other studies have used alerts to trigger sepsis workflows, including bedside assessment and order sets [11]. SCTP builds on these innovations by following patients suspected of having sepsis through the next several hours of care, which can be the most critical and prompts providers toward optimal care. Therefore, we expect it to influence provider action more directly for patients suspected of having sepsis and have a meaningful impact on clinical outcomes.

The focus of this study is on alert fatigue, as recent articles have shown that an increase in the number of alerts decreases the provider's acceptance rate [20] and many alerts in the EHR may not directly change clinical care [21]. The SCTP incorporated many design features to minimize alert fatigue, such as restricting enrollment to patients where the sepsis BPA was confirmed by an ED provider and condensing pages so no more than two pages are sent for any patient. Consultations with frontline providers and survey data were also used to optimize the message content to determine the appropriate page recipient. All of these factors likely contributed to the low number of pages sent per attending shift (0.98) and favorable survey responses. However, further work is needed to ensure that the clinical benefit of this system outweighs the alert fatigue it generates.

This study had several limitations. First, system implementation was conducted at a single academic site, which may not be generalizable across different health care settings. Second, the survey response rate was low for providers receiving alerts from the paging system. Subsequent iterations may consider text-based phone messaging or in-person surveys at the time of the provider receiving the page, which may encourage more detailed and complete responses. Third, the study was not designed to measure the effect size of SCTP paging alerts or review patient cases where providers understand sepsis to be unlikely despite the initial BPA. Our institution is currently initiating a randomized controlled trial to determine the efficacy of SCTP in improving compliance with SEP-1 bundle elements and patient outcomes. It will also further quantify any marginal increase in alert fatigue. Finally, there may be other simultaneous quality improvement initiatives and the Hawthorne effect that may confound our data.

Conclusions

In conclusion, we have demonstrated a feasible, technically sound, real-time provider alert system that has the potential to improve the SEP-1 compliance rate without significant amounts of alert fatigue. We anticipate that future work will involve a randomized controlled trial to measure potential increases in SEP-1 compliance rates and patient outcomes and further characterize alert fatigue. The expansion of this promising intervention could be considered in other inpatient settings.

Authors' Contributions

AHL, EA, KAH, and KCS designed the study, analyzed the data, and wrote the manuscript. MHF and HR adapted and implemented the electronic monitoring platform. EM and JDS assisted with carrying out the study in the hospital and editing the manuscript. All the authors provided critical feedback and contributed to the manuscript.

Conflicts of Interest

None declared.

References

1. Liu V, Escobar GJ, Greene JD, Soule J, Whippy A, Angus DC, et al. Hospital deaths in patients with sepsis from 2 independent cohorts. J Am Med Assoc 2014 Jul 02;312(1):90-92. [doi: 10.1001/jama.2014.5804] [Medline: 24838355]
2. Rhee C, Dantes R, Epstein L, Murphy DJ, Seymour CW, Iwashyna TJ, CDC Prevention Epicenter Program. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. J Am Med Assoc 2017 Oct 03;318(13):1241-1249 [FREE Full text] [doi: 10.1001/jama.2017.13836] [Medline: 28903154]
3. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to treatment and mortality during mandated emergency care for sepsis. N Engl J Med 2017 Jun 08;376(23):2235-2244 [FREE Full text] [doi: 10.1056/NEJMoa1703058] [Medline: 28528569]
4. Severe sepsis and septic shock: management bundle (Composite Measure). Centers for Medicare and Medicaid Services Measures Inventory Tool, Department of Health Human Services, USA. URL: https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=5175 [accessed 2019-09-17]
5. Mah JW, Bingham K, Dobkin ED, Malchiodi L, Russell A, Donahue S, et al. Mannequin simulation identifies common surgical intensive care unit teamwork errors long after introduction of sepsis guidelines. Simul Healthc 2009;4(4):193-199. [doi: 10.1097/sih.0b013e3181abe9d6]
6. Rhee C, Filbin MR, Massaro AF, Bulger AL, McEachern D, Tobin KA, et al. Compliance with the National SEP-1 quality measure and association with sepsis outcomes. Crit Care Med 2018;46(10):1585-1591. [doi: 10.1097/ccm.0000000000003261]
7. Hospital compare. Centers for Medicare & Medicaid Services. URL: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalCompare.html [accessed 2021-06-02]
8. Rhee C, Chiotos K, Cosgrove SE, Heil EL, Kadri SS, Kalil AC, et al. Infectious diseases society of America position paper: recommended revisions to the national severe sepsis and septic shock early management bundle (SEP-1) sepsis quality measure. Clin Infect Dis 2021 Feb 16;72(4):541-552. [doi: 10.1093/cid/ciaa059] [Medline: 32374861]
9. Rhee C, Gohil S, Klompas M. Regulatory mandates for sepsis care — reasons for caution. N Engl J Med 2014 May;370(18):1673-1676. [doi: 10.1056/nejmp1400276]
10. Aaronson EL, Filbin MR, Brown DF, Tobin K, Mort EA. New mandated centers for medicare and medicaid services requirements for sepsis reporting: caution from the field. J Emerg Med 2017 Jan;52(1):109-116. [doi: 10.1016/j.jemermed.2016.08.009] [Medline: 27720289]
11. Joshi M, Ashrafian H, Arora S, Khan S, Cooke G, Darzi A. Digital alerting and outcomes in patients with sepsis: systematic review and meta-analysis. J Med Internet Res 2019 Dec 20;21(12):e15166 [FREE Full text] [doi: 10.2196/15166] [Medline: 31859672]
12. Downing NL, Rolnick J, Poole SF, Hall E, Wessels AJ, Heidenreich P, et al. Electronic health record-based clinical decision support alert for severe sepsis: a randomised evaluation. BMJ Qual Saf 2019 Sep 14;28(9):762-768 [FREE Full text] [doi: 10.1136/bmjqs-2018-008765] [Medline: 30872387]
13. Narayanan N, Gross K, Pintens M, Fee C, MacDougall C. Effect of an electronic medical record alert for severe sepsis among ED patients. Am J Emerg Med 2016 Feb;34(2):185-188. [doi: 10.1016/j.ajem.2015.10.005] [Medline: 26573784]
14. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). J Am Med Assoc 2016 Feb 23;315(8):801-810. [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
15. Jones TW, Smith SE, Van Tuyl JS, Newsome AS. Sepsis with preexisting heart failure: management of confounding clinical features. J Intensive Care Med 2020 Jun 04:A. [doi: 10.1177/0885066620928299] [Medline: 32495686]
16. Truong TN, Dunn AS, McCardle K, Glasser A, Huprikar S, Poor H, et al. Adherence to fluid resuscitation guidelines and outcomes in patients with septic shock: reassessing the "one-size-fits-all" approach. J Crit Care 2019 Jun;51:94-98. [doi: 10.1016/j.jcrc.2019.02.006] [Medline: 30784983]
17. Seymour CW, Kahn JM, Martin-Gill C, Callaway CW, Yealy DM, Scales D, et al. Delays from first medical contact to antibiotic administration for sepsis*. Crit Care Med 2017;45(5):759-765. [doi: 10.1097/ccm.0000000000002264]
18. Makam AN, Nguyen OK, Auerbach AD. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: a systematic review. J Hosp Med 2015 Jun;10(6):396-402 [FREE Full text] [doi: 10.1002/jhm.2347] [Medline: 25758641]
19. Ruppel H, Liu V. To catch a killer: electronic sepsis alert tools reaching a fever pitch? BMJ Qual Saf 2019 Sep 23;28(9):693-696 [FREE Full text] [doi: 10.1136/bmjqs-2019-009463] [Medline: 31015377]
20. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R, with the HITEC Investigators. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak 2017 Apr 10;17(1):36 [FREE Full text] [doi: 10.1186/s12911-017-0430-8] [Medline: 28395667]
21. Todd B, Shinthia N, Nierenberg L, Mansour L, Miller M, Otero R. Impact of electronic medical record alerts on emergency physician workflow and medical management. J Emerg Med 2021 Mar;60(3):390-395. [doi: 10.1016/j.jemermed.2020.10.017] [Medline: 33298357]

Abbreviations

BPA: best practice alert
CMS: Centers for Medicare and Medicaid Services

ED: emergency department
EHR: electronic health record
EMR: electronic medical record
SCTP: Sepsis Care Tracking Platform
SEP-1: sepsis core measure

Edited by R Kukafka; submitted 04.01.21; peer-reviewed by M Görges, JC Ruiz-Rodríguez; comments to author 01.02.21; revised version received 14.02.21; accepted 30.04.21; published 24.06.21.

Please cite as:
Lee AHY, Aaronson E, Hibbert KA, Flynn MH, Rutkey H, Mort E, Sonis JD, Safavi KC
Design and Implementation of a Real-time Monitoring Platform for Optimal Sepsis Care in an Emergency Department: Observational Cohort Study
J Med Internet Res 2021;23(6):e26946
URL: https://www.jmir.org/2021/6/e26946/
doi:10.2196/26946
PMID:34185009

Original Paper

Secondary Use of Clinical Data in Data-Gathering, Non-Interventional Research or Learning Activities: Definition, Types, and a Framework for Risk Assessment

Martin Jungkunz[1], Dr; Anja Köngeter[1], MA; Katja Mehlis[1], Dr; Eva C Winkler[1], Prof., MD, PhD; Christoph Schickhardt[2], Dr

[1]Section for Translational Medical Ethics, Department of Medical Oncology, National Center for Tumor Diseases, Heidelberg University Hospital, Heidelberg, Germany

[2]Section for Translational Medical Ethics, National Center for Tumor Diseases, German Cancer Research Center (DKFZ), Heidelberg, Germany

Corresponding Author:
Martin Jungkunz, Dr
Section for Translational Medical Ethics, Department of Medical Oncology
National Center for Tumor Diseases
Heidelberg University Hospital
Im Neuenheimer Feld 460
Heidelberg, 69120
Germany
Phone: 49 6221 56 35059
Email: martin.jungkunz@med.uni-heidelberg.de

Abstract

Background: The *secondary use of clinical data in data-gathering, non-interventional research or learning activities* (*SeConts*) has great potential for scientific progress and health care improvement. At the same time, it poses relevant risks for the privacy and informational self-determination of patients whose data are used.

Objective: Since the current literature lacks a tailored framework for risk assessment in *SeConts* as well as a clarification of the concept and practical scope of *SeConts*, we aim to fill this gap.

Methods: In this study, we analyze each element of the concept of *SeConts* to provide a synthetic definition, investigate the practical relevance and scope of *SeConts* through a literature review, and operationalize the widespread definition of risk (as a harmful event of a certain magnitude that occurs with a certain probability) to conduct a tailored analysis of privacy risk factors typically implied in *SeConts*.

Results: We offer a conceptual clarification and definition of *SeConts* and provide a list of types of research and learning activities that can be subsumed under the definition of *SeConts*. We also offer a proposal for the classification of *SeConts* types into the categories *non-interventional (observational) clinical research*, *quality control and improvement*, or *public health research*. In addition, we provide a list of risk factors that determine the probability or magnitude of harm implied in *SeConts*. The risk factors provide a framework for assessing the privacy-related risks for patients implied in *SeConts*. We illustrate the use of risk assessment by applying it to a concrete example.

Conclusions: In the future, research ethics committees and data use and access committees will be able to rely on and apply the framework offered here when reviewing projects of secondary use of clinical data for learning and research purposes.

(J Med Internet Res 2021;23(6):e26631) doi:10.2196/26631

KEYWORDS

secondary use; risk assessment; clinical data; ethics; risk factors; risks; privacy; electronic health records; research; patient data

Introduction

The secondary use of clinical data for research purposes is increasingly recognized as a promising and crucial tool for improving health care and advancing medical research. Several initiatives strive to use data from medical care for secondary research and learning activities [1]. The US Institute of Medicine has called for a transformation toward a learning health care system (LHCS) to improve quality, expedite translation, and reduce costs [2]. The American Society of Clinical Oncology is pioneering the linkage of patient, provider, and research data with their quality improvement portal CancerLinQ [3]. In Germany, the national Medical Informatics Initiative aims to establish a national network of data integration centers to collect, aggregate, and analyze clinical data from all university hospitals in Germany.

There are numerous advantages of the secondary use of clinical data (ie, data derived from patient care) for research and learning activities. First, the clinical data are readily available. There is no need for any physical intervention or data collection through surveys. Costs for data aggregation, staff, and materials are considered to be low; data can be collected quickly [4]. Data have a high level of generalization due to the real-life setting in which they are collected, and this facilitates representative sampling while simultaneously increasing external validity [5,6]. Moreover, large sample sizes can be obtained by aggregating the data from different sites. For example, this benefits research on rare diseases [7]. Moreover, interventional studies that cannot be conducted prospectively due to ethical reasons may be performed retroactively [8], for instance, by systematically analyzing experimental therapies such as off-label use of drugs. Most importantly, patients can contribute their clinical data to research or learning activities without being exposed to immediate physical risks [5]. However, the secondary use of clinical data in research or learning activities entails data-associated risks that require further investigation.

This paper focuses on the most relevant aspects for patients in particular and for the trustworthiness and sustainability of secondary use of clinical data in general: risks concerning patients' privacy and informational self-determination. The European Union General Data Protection Regulation (GDPR) requires data processors to carry out an "assessment of the impact of the envisaged processing operations on the protection of personal data" where there is a high risk to the "rights and freedoms of natural persons" (Article 35, 1, GDPR). An appropriate framework for risk assessment of the secondary use of clinical data in research or learning activities is lacking, as is a conceptual basis for such secondary data use. We aim to fill this gap by developing a framework for risk assessment that supports decision makers in research ethics committees and data use and access committees, as well as scientists, bioethicists, and funders who deal with the ethics and governance of secondary use of clinical data in data-gathering, non-interventional research or learning activities (ie, research and learning activities that rely solely on the collection of existing data). We will develop the risk assessment framework on a strong conceptual and empirical basis of two preceding elementary steps: (1) an analytical clarification of the concept of secondary use of clinical data in data-gathering, non-interventional research or learning activities (*SeConts*); and (2) an illustration of the types of research and learning activities that can be subsumed under this concept.

Methods

In this study, we proceed in three main steps, each with a methodological approach. In *step one*, our methodological approach to clarify the concept of *SeConts* is to investigate the *intension*. In other words, what does the concept of *SeConts* mean? We first analyze each element of the concept (analysis) and then compile them for a comprehensive definition of the concept (synthesis). As the application of this methodological approach to clarify the concept cannot easily be separated from the result itself (the definitional clarification of the concept), we decided not to separate them and thus present both parts in the *Results* section.

Building on the clarification of the concept of *SeConts* (step 1), *step 2* examines its practical relevance. Regarding our methodological approach, we investigate the *extension* of the concept of *SeConts*, that is, the range of *objects* to which the concept can be applied. We examine concrete types of research or learning activities that can be subsumed under this concept. These types of research or learning activities were inferred from a scoping review (a), in which we searched PubMed and Google Scholar between October and November 2019 for bioethical literature that deals with the expected risks and benefits of secondary use of clinical data for biomedical research. The search terms used were *ethics*, *secondary use*, *re-use*, *clinical data*, *electronic health records*, *risks*, and *benefits* in different combinations. We limited our search to publications in English. We found numerous types of research or learning activities that, according to the authors, solely used clinical data. Next, we searched the biomedical literature for concrete studies (b) to find examples of previously identified types of research or learning activities. Relevant publications were identified after reading abstracts. After reading the full texts of the identified publications, we categorized them into the types of *SeConts* developed before (a).

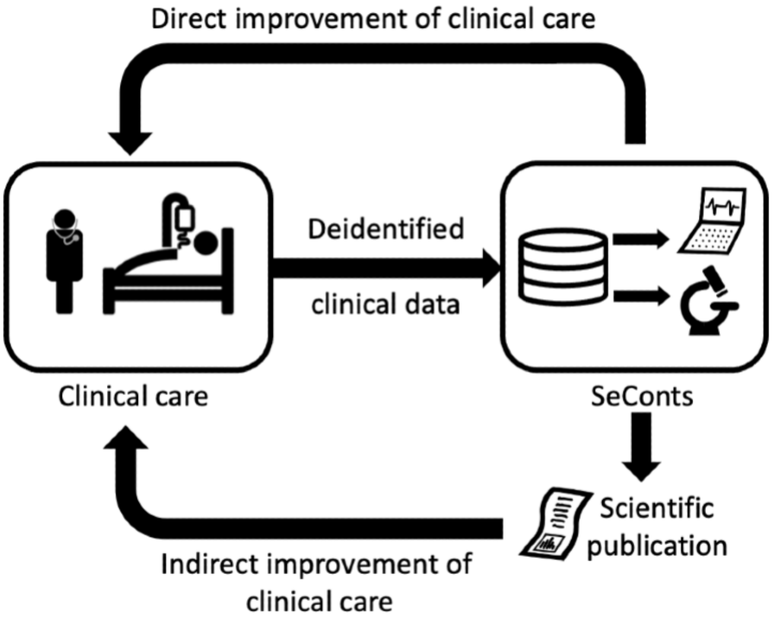
On the basis of step 1 (conceptual clarification of *SeConts*) and step 2 (examination of the practical relevance of *SeConts*), we develop a framework for the systematic assessment of risks implied in *SeConts* in *step 3*. Operationalizing the general concept of risk to tailor it to *SeConts* presents a key methodological challenge when developing a risk assessment framework. We apply a widespread definition of risk as a *harmful event of a certain magnitude* that occurs with a certain *probability* [9-11]. According to this definition, risk assessment ideally results in a number. If we bet US $10 on a single number out of 37 in the roulette, the probability of losing is $1-(1/37)$ and the magnitude of harm is US $10. The resulting risk could therefore be quantified by the number of 9.7, which is $10\times[1-(1/37)]$. Although this quantitative understanding of risk is intuitively plausible, it is fraught with several difficulties in the context of *SeConts* concerning the quantification of *the probability* and *magnitude of harm*.

In terms of quantifying the *probability* of a harming event in the context of *SeConts*, there is neither a stochastic rule (as in the roulette example) nor empirical data that would allow an estimate of the probability of these events. Reports indicate that there have been data breaches (defined here as: *all types of events in which the confidentiality and planned protection of data is violated, whether through technical failure, human error, or deliberate unethical or criminal acts*) in the health care system in the past [12-14] that may suggest that such events could also occur in the context of *SeConts*. However, they do not allow for an assessment of their likelihood. Similarly, there is little information about the probability of further misuse of data stolen in the course of a breach. The *magnitude of harm* caused by the misuse of data is equally difficult to quantify. This is due to the fact that possible harm is predominantly of an immaterial social and psychological nature.

In the field of data science and in bioethics, various approaches have been developed to address the problem of operationalizing risks without recourse to quantification [15-20]. However, these approaches are neither designed nor transferable to *SeConts*. In this paper, we therefore propose a nonquantitative approach to operationalize both the *probability* and *magnitude of harm* for the assessment of risks within *SeConts*. We (1) identify risk factors in relation to data and their use within *SeConts* that increase or decrease the probability and magnitude of harm, (2) assess their individual severity, and (3) subsequently estimate the overall risk of a specific form of *SeConts*.

Similar to existing initiatives for secondary use of clinical data, such as the Medical Informatics Initiative [21] or CancerLinQ [3], we assume the following data flow from patients to *SeConts* (Figure 1): (1) Clinical data from electronic health records (EHRs) are deidentified, that is, identifying attributes (eg, names, addresses) are removed or replaced with a code. (2) Deidentified data are transferred to a central data warehouse and stored. (3) Deidentified data stored in the central data warehouse can be made available to researchers upon request to enable *SeConts*. Regarding risks to the confidentiality of EHRs that contain directly identifying data, there is evidence of leaks and attacks on EHRs. However, these risks affect every care unit that works with EHRs and are not specific to *SeConts*. Therefore, the risks to the confidentiality of EHRs are not the focus of our study.

Figure 1. Flowchart of the secondary use of clinical data in data-gathering, non-interventional research or learning activities.



Regarding the operationalization of *probability*, we assume that the most negative consequences for the individual patient are only expected if the patient can be identified from the data. Therefore, we assume unauthorized reidentification as a prerequisite for possible harm in connection with *SeConts*. According to our understanding, factors that increase the *probability of unauthorized reidentification* also increase the probability of misuse. We identify two categories of factors that directly or indirectly determine the *probability of unauthorized reidentification*: data-specific and contextual factors.

In terms of operationalizing the *magnitudeof harm*, we distinguish different types of data according to their harm potential (see studies by Dyke at al [20] and Rumbold and Pierscionek [22]). To do this, we identify risk factors in the data that determine the extent of harm that may result from the misuse of the data. In other words, we provide an analytical answer to the question of what data aspects determine the degree of harm in the event of misuse. This approach is in line with the risk-based distinction between different categories of data in the GDPR (Article 9 GDPR).

How did we arrive at the factors that determine the *probability of unauthorized reidentification* and *magnitude of harm*? In the *first step*, we built on existing literature. As a basis for the factors determining the *probability of unauthorized reidentification*, we focused on the literature from different areas: literature on ethical, legal, and social implications of *SeConts*; literature on data security, data protection, and the assessment of reidentification risks; and national German and European data protection laws and regulations. For factors that determine the *magnitude of harm*, we analyzed the literature on genetic data that are generally classified as very sensitive, to learn what makes these data sensitive and transfer this knowledge, mutatis mutandis, on other types of data. In the *second step*, we extracted the individual factors—for both the *probability of unauthorized reidentification* and the *magnitude of harm*—from the literature and categorized them inductively. In the *third step*, we cross-referenced the factors with qualitative interviews conducted with experts from research, care, medical informatics, patient advocacy, and politics on the topic of perceived risk potentials of *SeConts*, part of which are to be published elsewhere [23]. In the presentation of the individual factors below, we refer to the literature on which they are based. Factors without a literature reference are taken from expert interviews that have not yet been published.

Results

Conceptual Clarifications and Definition of Secondary Use of Clinical Data in Data-Gathering, Non-Interventional Research or Learning Activities

Secondary use of health data is defined by the American Medical Informatics Association as "non-direct care use of PHI [personal health information] including but not limited to analysis, research, quality/safety measurement, public health, payment, provider certification or accreditation, and marketing and other business including strictly commercial activities" [24]. On the basis of a systematic scoping review, Robertson and colleagues categorized secondary use by distinguishing between four types of secondary use of clinical data: research, improving quality and safety of care, informing financial management, and education [25]. These and other similar classifications [26-28] help in understanding the broad spectrum of secondary uses of clinical data and illustrate that it is not limited to research. However, as pertains to the focus of this paper, that is, *the secondary use of clinical data in data-gathering, non-interventional research or learning activities*, the generic classifications lack further specifications and detail.

The term *secondary use* (or reuse) implies that there is also a primary use. Primary use encompasses the generation and use of data within the context of individual health care in hospitals and doctors' offices to serve direct care needs. Secondary use refers to the use of these data for purposes other than individual care. Two points of criticism could be made against the distinction between primary and secondary use and the implied distinction between care and research. First, the distinction between care and research is notoriously difficult and widely discussed. Second, if secondary use of clinical data was to be implemented in the future as a standard in the health care system (ie, all health care data would be made available for possible research), this could lead to a tendency to collect data more systematically or collect more data than necessary in the care context, thereby blurring the distinction between data generation for care (primary use) and data generation for research (secondary use). In arguing against these two criticisms, however, we find that (1) ethical and legal codes are still fundamentally based on the distinction between care and research, which is unlikely to be abandoned in the coming years as blurred boundaries in governance and regulation are difficult to manage. Furthermore, we argue (2) that even if data are collected with the additional motivation of secondary use, the generation and collection of data would be driven by the primary goal of individual care.

Clinical data refers to data generated and collected in clinical contexts for patient care (diagnoses, anamnesis, treatment, medication, and so on) as well as data for accounting purposes or patient management, such as age, employment status, and other sociodemographic information. Clinical data also include data generated in the course of nonroutine treatment (off-label or experimental therapies) as long as it is collected for the primary objective of individual care. In terms of clinical data, we also include data reported from health care services and units to health insurance. In contrast to the widely used term health data, clinical data includes neither data gathered by (common, ie, nonclinical) smart devices and smartphone apps or research data understood as data generated for research purposes.

In contrast to research that generates data, *SeConts* only *gathers existing clinical data*. The gathering of data in *SeConts* can be done either by collecting clinical data of different patients from one institution or several institutions or by collecting only the results of analyses of clinical data carried out in the institution of primary use (decentralized analysis).

SeConts is *non-interventional* as it does not use data from interventions carried out with the aim or priority of data generation for research. *SeConts* solely uses existing data from medical care. The data used in *SeConts* may originate from interventions, but only from interventions carried out for the sake of individual diagnosis and treatment.

Both the terms *research* and *learning activities* refer to investigations in which the acquisition of generalizable knowledge beyond the needs and logic of individual care is the sole or primary intension. Research and learning activities cannot be clearly separated, but a rough distinction can be made in terms of their respective objectives. Research primarily aims at acquiring scientifically generalizable knowledge to be shared within the scientific community through scientific publications. Research usually indirectly improves health care and is realized through publications and by implementing practical conclusions based on research results. Learning activities, on the other hand, are designed to acquire knowledge about current care practices (eg, in a defined care unit) to derive appropriate and immediate measures to directly improve a specific health care unit or service. When talking about improving health care, the question arises whether *SeConts* has the potential to directly benefit the individual patient whose clinical data are reused. A direct (therapeutic) benefit is only possible under certain circumstances, for example, infection control of a clinical unit or research on a chronic disease. Younger patients with such a chronic disease could possibly benefit in several years from research that improves treatment of that very disease. However, as such benefits seem rather unlikely and exceptional, we understand *SeConts* as activities that are neither intended nor expected to directly benefit the individual patient whose data are used. In addition, there are ethical reasons for this narrow definition of *SeConts*, namely, to avoid any possibility of therapeutic misunderstanding.

In summary, the central concept of this paper, that is, *SeConts* can be defined as activities that:

1. exclusively use data produced for the purposes of and in the context of health care.
2. exclusively collect and do not generate data, that is, they are not based on data generated by interventions carried out primarily for the sake of research.

3. aim to acquire generalizable knowledge that goes beyond the needs and logic of individual care.
4. aim to directly improve health care units or services or publish their results for the promotion of biomedical science.

In a nutshell, *SeConts* describes *activities that solely use data produced for the sake of health care and in the context of health care to improve biomedical science or services.*

Overview of Different Types of Secondary Use of Clinical Data in Data-Gathering, Non-Interventional Research or Learning Activities

Having clarified the concept of *SeConts* in the previous section, we now examine the scope and practical relevance of the concept of *SeConts* as previously defined. Which studies fall under the concept of *SeConts* (scope)? What is their practical relevance to medical research and improving medical care? We explore these questions in the next section. The literature review carried out to investigate the scope and practical relevance of research or learning activities that fall within our definition of *SeConts* (step 1) led to the following results: many types of research or learning activities common in the field of quality control and quality improvement can be subsumed under the concept of *SeConts* (in particular, under the term *learning activities*). Examples include *improvement of infection control*, which can be done by monitoring clinical data in hospitals to identify patients at high risk of infection [29]. Clinical data can also be used to create computerized algorithms for the *early detection of possible hazards from germs* [30]. These activities can be considered a component of the comprehensive ideal for transforming a particular health care institution into an LHCS.

At the national level, clinical data are used for *public health surveillance*. For instance, data from EHRs are searched for indicators of influenza in primary care to detect a pandemic in its early phase [31]. *Epidemiological studies* focus on the distribution of diseases as well as their causes and effects in populations, such as studies on the epidemiology of a certain infection to inform and improve vaccination initiatives [32]. In *outcomes research*, the effects (outcome) of certain interventions are investigated, such as the effects of a nationwide antismoking law on childbirth in the area of public health [33] or, on a clinical care level, evaluating the quality of care [34]. In *health services research*, investigators can use clinical data to explore the mean costs associated with (treatment of) a certain disease [35]. A well-established form of secondary use of clinical data is *registry studies* analyzing collections of data on all patients affected by a particular disease (registries) such as cancer registries [36].

Clinical data are also reused in clinical research in the form of *in-silico hypothesis testing*, where clinical trials are modeled with the help of data from EHRs [37]. Moreover, clinical data can be used for *comparative effectiveness research* [38] to compare “the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care” [39]. *Evaluation of experimental therapies* can be conducted in terms of *SeConts*, for instance, by sharing data from single off-label (or compassionate) use from different hospitals. The secondary use of data from these therapies can help inform other physicians with similar patients about the course and outcome of different experimental therapeutic approaches [40]. Other studies that reuse clinical data in the sense of *SeConts* are *drug safety and efficacy studies* [41]. In addition, some basic research is conceivable using existing clinical data, such as *studies on risk factors* for certain diseases that can be linked to single influencing or moderating factors [42,43]. Another form of secondary use of clinical data lies in the area of *informatics research*, which uses clinical data to develop new software tools that have the potential to improve patient care or analyze and improve data security within a health system [44].

In addition to the different studies mentioned above, clinical data can be used in an *explorative* manner, which can be understood as encompassing three (potentially subsequent) steps. First, data can be analyzed (possibly using artificial intelligence) to *generate hypotheses* for future studies; second, in the case of new research ideas resulting from this, the data can be used to *check feasibility*; and third, to *identify potential participants* for recruitment of upcoming studies [45].

To further clarify the scope and concept of *SeConts* (by means of definitio ex negativo), we also want to mention two areas beyond the concept of *SeConts*: prospective clinical trials (eg, randomized controlled trials) and all other forms of studies that include interventions carried out for the sake of generating data; genome wide association studies in so far as they combine clinical data (phenotypes) with sequencing data generated for research purposes (genotypes).

Thus far, we have illustrated that there are many types of research or learning activities that can be conducted in the sense of *SeConts*. These types of research or learning activities display huge heterogeneity with respect to study designs, research questions, and contexts. Even the names of the types show this heterogeneity, as they refer to very different attributes describing the respective type of *SeConts*: some names refer to a specific method or study design (eg, in-silico hypothesis testing), whereas others refer to a subject area (eg, epidemiology). Given this heterogeneity, the question arises as to whether the different types of *SeConts* can be categorized in a way that allows for a better overview. The literature that attempts to capture the potential applications of secondary use of clinical data does not systematically distinguish between different types of research and learning activities [25,26]. Therefore, we searched for criteria or a particular logic that would provide some kind of categorization of the different types of *SeConts*. Some criteria that provide useful approaches to classification in other biomedical contexts are inappropriate. For example, although categorization by medical specialty seems conceivable at first glance, we determined it to be unhelpful on closer inspection. With all the different specialties that exist in medicine, such a categorization does not provide any reduction of heterogeneity. Categorization into common typologies of biomedicine, such as *basic research*, *clinical research*, and *translational research*, is also not useful, because learning activities, as described above, cannot be clearly assigned here. Ultimately, we found a categorization in terms of the *object of investigation* (ie, the entity about which *SeConts* seeks to produce knowledge) using three levels most appropriate. These three levels are familiar from the social sciences: micro level, meso level, and macro level. The object of investigation can be either patients (micro level), clinical care units such as clinical departments or single hospitals—a specific delimited patient population (meso level), or the general public (macro level). Taking the *object of investigation* as a criterion, we can attribute each type of research or learning activities to a general area of application: *non-interventional (observational) clinical research* focusing on individuals, *quality control and improvement* (and similar uses contributing to the creation of an LHCS) focusing on clinical units, and *public health research* focusing on the general public. Table 1 sums up all types of the aforementioned research or learning activities, including the object of investigation and area of application. It also shows that some types of research or learning activities are not limited to a single object of investigation or area of application.

Table 1. Possible types of research or learning activities within secondary use of clinical data in data-gathering, non-interventional research or learning activities.

Type of research or learning activities	Object of investigation	Area of application
Improvement of infection control	Clinical unit	Quality control and improvement
Early detection of possible hazards from germs	Clinical unit	Quality control and improvement
Public health surveillance	General public	Public health research
Epidemiology	General public	Public health research
Outcomes research	Patients or clinical unit	Public health research or quality control and improvement
Health services research	General public	Public health research
Register studies	General public	Public health research
In-silico hypothesis testing	Patients	Non-interventional (observational) clinical research
Comparative effectiveness research	Patients	Non-interventional (observational) clinical research
Experimental therapy evaluation	Patients	Non-interventional (observational) clinical research
Drug safety and efficacy studies	Patients	Non-interventional (observational) clinical research
Studies on risk factors	Patients	Non-interventional (observational) clinical research
Medical informatics research	Patients, clinical unit, or general public	Possible in all three areas of application
Explorative use	Patients, clinical unit, or general public	Possible in all three areas of application

Tailored Framework for Risk Assessment of Secondary Use of Clinical Data in Data-Gathering, Non-Interventional Research or Learning Activities

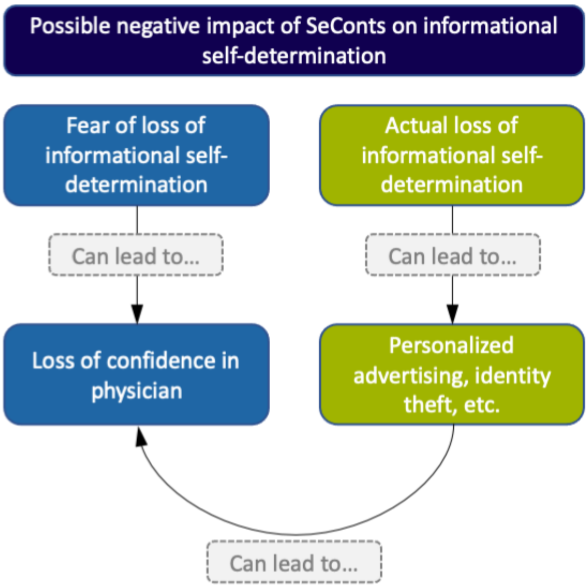
Risks for Informational Self-determination and Confidentiality

In the previous section, we addressed two desiderata of the current literature on *SeConts* by defining our understanding of *SeConts* and clarifying its scope and practical relevance. These steps allow us to offer what we consider to be the third desideratum: a detailed analysis of risks for patients implied in *SeConts* as well as a framework to assess these risks. The risks to patients associated with the data are critical to *SeConts* because if they were to materialize, they could also have a major impact on the role, trustworthiness, and reputation of doctors, hospitals, and the public health system. *SeConts* implies potential risks for patients because of their (potentially reidentifiable and sensitive) clinical data being used. We assume that *SeConts* will usually rely on clinical data that cannot be classified as anonymous data. Even if direct identifiers are removed from data or replaced by a code (deidentification), the possibility of reidentifying a specific person in the data can rarely be excluded (see the list of potential risk factors below). In most cases, *SeConts* will use deidentified personal (or person-related) data; they imply risks for patients' right to *informational self-determination*.

The term informational self-determination refers to "a person's ability to freely decide whether and how personal data and information about her are collected, stored, multiplied, processed, and transferred by third parties" [46]. We regard informational self-determination as an instrumental value, that is, a means of protecting liberal individual and social rights and values such as equality or personal autonomy [47-49]. The protection of informational self-determination is particularly important in the health care sector: clinical data are highly sensitive as they contain information about the health status of a person and can therefore be highly susceptible to possible misuse. The confidentiality of the data is crucial for the patient-physician relationship and communication.

SeConts can have a negative impact on the patient's ability to exercise informational self-determination in two ways (Figure 2): (1) the fear of a loss of informational self-determination and (2) its actual loss in case of data breaches in the meaning defined above (which, in turn, will most likely reinforce the fear of a loss of informational self-determination among patients).

Figure 2. Possible negative impact of the secondary use of clinical data in data-gathering, non-interventional research or learning activities on the patient's ability to exercise informational self-determination.



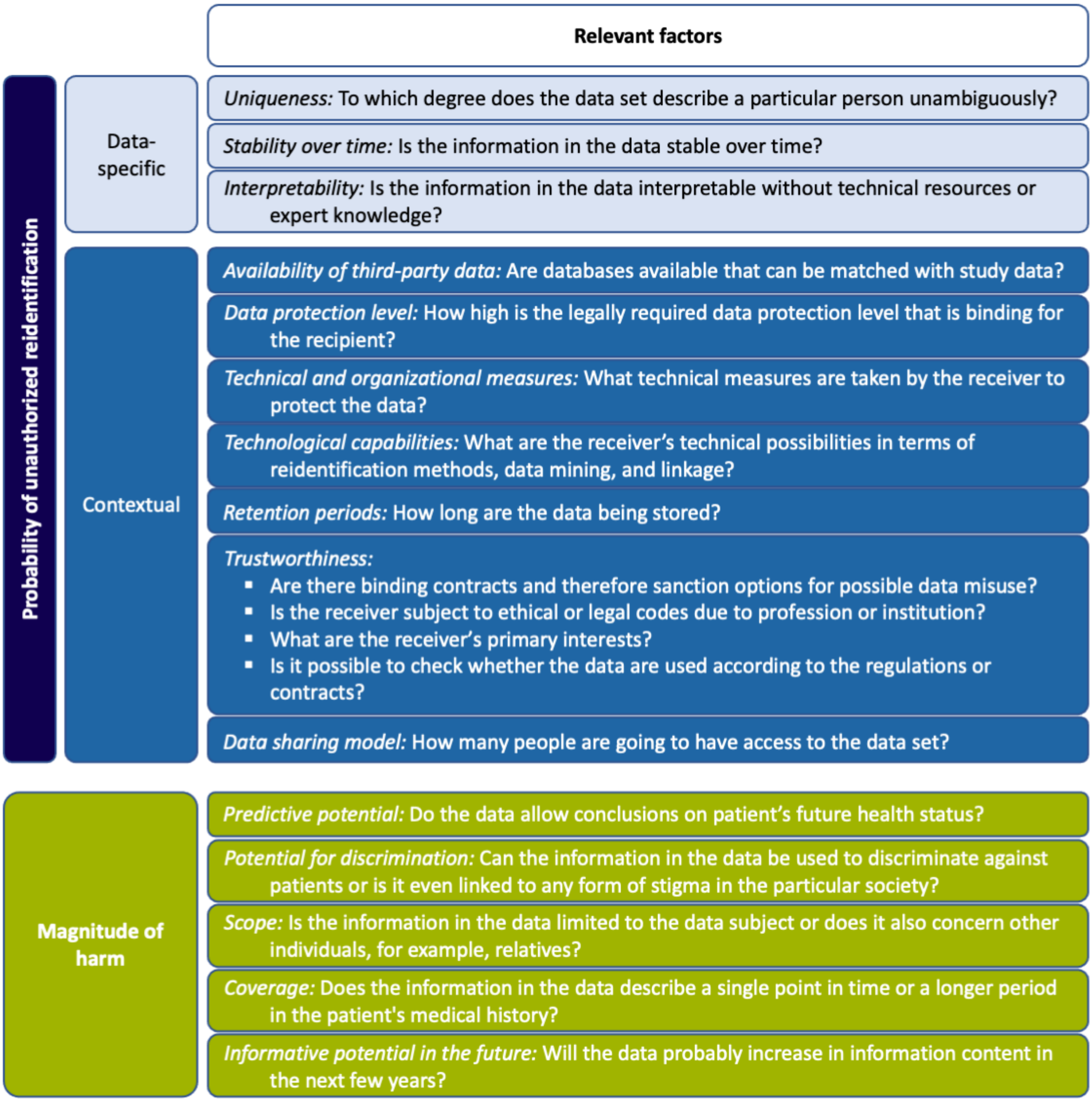
With regard to fear (1), it is important to stress that informational self-determination is not only impaired if there are indications that third parties have accessed the data without authorization (and might or do use them against the data subject). Informational self-determination is already undermined if data subjects have reasons to develop a sense of vague uncertainty about the confidentiality of personal data [50] (chilling effect [51]). The fear of loss of informational self-determination alone can already have negative consequences. For instance, the fear that their data are not protected from unauthorized access can lead to patients not disclosing all the information that could be important for their personal care [52,53] due to a lack of confidence in their physician.

Concerning actual losses of informational self-determination (2), data breaches are a serious threat. A study carried out between 2010 and 2013 revealed a total of 949 data breaches in American hospitals involving almost 30 million patient records [14]. Even if the reported data breaches occurred in the context of patient care, and not in the context of *SeConts*, they uncover general problems with the protection of clinical data. As assumed by several authors [52,54,55], the negative consequences of data breaches, including unauthorized reidentification, could range from minor annoyances through personalized advertising to serious harms such as identity theft, stigmatization, blackmail, or discrimination, as well as other forms of data use without patient consent, such as the sale of data or use in studies in which patients have not consented.

Criteria to Assess Risk: Probability of Unauthorized Reidentification and Magnitude of Harm

In the previous chapter, we outlined the potential risks to patients' informational self-determination associated with *SeConts*. On the basis of our approach to assessing these risks (probability and magnitude of harm, see *Methods* section), we present a list of relevant factors that determine the *probability of unauthorized reidentification* and the *magnitude of harm*. These factors are partly interrelated and cannot always be clearly distinguished. Figure 3 provides an overview of the relevant factors, which we discuss in more detail below.

Figure 3. Factors of risk assessment for the secondary use of clinical data in data-gathering, non-interventional research or learning activities.



Relevant factors

Probability of unauthorized reidentification

Data-specific

Uniqueness: To which degree does the data set describe a particular person unambiguously?

Stability over time: Is the information in the data stable over time?

Interpretability: Is the information in the data interpretable without technical resources or expert knowledge?

Contextual

Availability of third-party data: Are databases available that can be matched with study data?

Data protection level: How high is the legally required data protection level that is binding for the recipient?

Technical and organizational measures: What technical measures are taken by the receiver to protect the data?

Technological capabilities: What are the receiver's technical possibilities in terms of reidentification methods, data mining, and linkage?

Retention periods: How long are the data being stored?

Trustworthiness:

- Are there binding contracts and therefore sanction options for possible data misuse?
- Is the receiver subject to ethical or legal codes due to profession or institution?
- What are the receiver's primary interests?
- Is it possible to check whether the data are used according to the regulations or contracts?

Data sharing model: How many people are going to have access to the data set?

Magnitude of harm

Predictive potential: Do the data allow conclusions on patient's future health status?

Potential for discrimination: Can the information in the data be used to discriminate against patients or is it even linked to any form of stigma in the particular society?

Scope: Is the information in the data limited to the data subject or does it also concern other individuals, for example, relatives?

Coverage: Does the information in the data describe a single point in time or a longer period in the patient's medical history?

Informative potential in the future: Will the data probably increase in information content in the next few years?

There are three data-specific factors that determine the probability of unauthorized reidentification:

1. *Uniqueness* of a data set [56,57]: Even after removing identifying attributes (name, date of birth, etc), unauthorized reidentification is possible, as the combination of attributes of a person's medical record can be identifying. Therefore, the more unambiguously a person is represented in a data set by the combination of individual attributes, the higher the probability of unauthorized reidentification. Statistically, the smaller the number of cases in a data set, the more likely it is to correctly assign one data set to a specific individual. In addition, a larger number of variables (attributes of patients) in a data set allows for a more unique combination of attributes that again increases the probability of unauthorized reidentification [56,58,59]. Concepts such as *k*-anonymity [60], *l*-diversity [61], and*t*-closeness [62] aim to mitigate these risks by defining standards for data sets to reduce the probability of unauthorized reidentification.
2. *Stability over time* [63]: The more stable the attributes described in the data, the more likely it is to link the data set to a specific person. If an attribute described in the data is not permanent and is likely to change between each measurement (eg, blood pressure or blood glucose levels), it is impossible to use the associated data to uniquely assign it to an individual patient.
3. *Interpretability:* Certain types of data require special skills or technical tools to understand them. Thus, the degree to which data is easier or more difficult to interpret determines the number of people who are able to misuse the data. For example, anyone can interpret data relating to obesity, but laypersons are currently unable to interpret data from genetic sequencing or magnetic resonance imaging.

However, the interpretability of data is a dynamic factor. Modern software development suggests that the interpretation of certain data might soon be supported by self-learning algorithms that may allow laypersons to interpret complex data such as genomic data or magnetic resonance imaging. *Interpretability* is directly related to the possible contexts in which data are used.

In addition to the data-specific factors regarding the probability of unauthorized reidentification in the context of *SeConts*, seven contextual factors must be considered.

1. *Third-party data:* Personal data of patients stored by third parties, such as voting registers or records from residents' registration offices, can be used for reidentification when linked to clinical data [64-66]. Therefore, the availability of third-party data makes unauthorized reidentification easier [63], especially when the factor of uniqueness of the data set is high.
2. *Data protection level:* Every act of data exchange with other institutions or countries or jurisdictions brings with it the possibility of weaker data protection standards. The legally required data protection level that is binding to the recipient must therefore be taken into account [67].
3. *Technical and organizational measures:* Adequate measures need to be taken by the data recipient to protect the data. Here, the data protection level of the recipient, for example, according to approved codes of conduct (Article 40, EU-GDPR) or a data protection certification (Article 42, EU-GDPR), can be used as a benchmark.
4. *Technological capabilities:* The factor *technological capabilities* describes the technical means available to the respective data recipient to perform unauthorized reidentification. These include self-learning algorithms or other forms of artificial intelligence [68].
5. *Retention periods*: The longer the data are stored, the higher the risk of a data breach that can lead to unauthorized reidentification. Therefore, the retention (and deletion) periods specified by the data recipient are relevant factors in estimating the probability of unauthorized reidentification.
6. *Trustworthiness:* The factor *trustworthiness* is determined by several aspects [67]: the existence of binding (sanctionable) contracts that determine the terms of data use and access; the existence of ethical or legal codes on the part of the data recipient [22]; the primary interest of the receiver, for example, monetary or public interests; the verifiability of the use of the data in accordance with the applicable regulations or contracts.
7. *Data sharing model:* Depending on the data sharing model, data are disclosed to a different number of people; for example, data can be disclosed only to a small work group, a consortium, or can be uploaded to a public database. Every person with access to the data increases the probability of unauthorized reidentification.

Together with the aforementioned factors determining the probability of unauthorized reidentification, the following five factors must be considered to determine the magnitude of possible harm.

1. *Predictive potential:* The predictive potential of data refers to the extent to which a data set contains information that allows insights regarding future health status [69-71]. If a data set has a time-limited relevance, it can be considered less problematic than a data set that allows conclusions to be drawn about a person's expected state of health for the next 30 years. Information with predictive potential could be, for example, the diagnosis of a chronic disease, but also the documentation of traumatic events that make a posttraumatic stress disorder diagnosis very likely within the next few years. In contrast, information about a fracture, for instance, does not necessarily provide information about future health status.
2. *Potential for discrimination:* Some data have obvious potential for discrimination because they contain information that can be used against individuals and groups. For example, it is conceivable that some employers discriminate against employees (eg, by not renewing contracts) if they know that the employee is affected by an illness (or has a disposition to illness) that is likely to lead to a longer absence of the employee in the future. Furthermore, stigmatization is possible [20]. For example, the disclosure of an HIV diagnosis can lead to stigmatization (through prejudice or social rejection).
3. *Scope:* The factor *scope* describes whether the information, and thus the possible risk in the data is limited to the data subject or whether it allows inferences about family members (eg, in the case of genomic information [69]) or colleagues (eg, information on diseases related to certain working conditions) [70].
4. *Coverage:* The factor *coverage* describes the data in terms of the period in which they were collected. Data from a single hospital visit cover a less extensive period than data documenting the entire medical history, for example, as collected by insurance companies. A complete medical history is likely to provide a more multifaceted picture of a person's life than data collected on a single point in time and can therefore potentially cause greater harm in the event of misuse.
5. *Informative potential in the future:* Some data types can be expected to contain more information than can currently be made available. For example, we currently assume that the information content of genomic data is still largely unknown [70]. However, the constant progress in the field of genomic research suggests that in the future, we will be able to retrieve significantly more information from genomic data than is the case today, which can potentially cause greater harm in case of misuse.

Discussion

Application and Use of the Risk Assessment

After introducing the individual factors that determine the *probability of unauthorized reidentification* and the *magnitude of harm* implied by *SeConts*, questions arise concerning their application and use. How do these factors form a comprehensive framework for risk assessment in the context of *SeConts*? How can the framework be applied? How can the assessment results be used? Who may apply and use this framework?

With our risk assessment, we intend to support the analysis, evaluation, and potential decision-making process of research ethics committees and data use and access committees, as well as scientists, bioethicists, and funders investigating the ethical acceptability of requests for concrete types of *SeConts*. We recommend the following procedure with five consecutive steps to apply and use the risk assessment framework.

Application of the Risk Assessment (Steps 1 and 2)

The first two steps concern the application of the risk assessment framework:

- Step 1—*identification and evaluation of single risk factors:* A concrete projected study is examined in light of all factors determining the probability of unauthorized reidentification and magnitude of harm listed in Figure 3. The individual severity of each factor is evaluated, that is, whether and to what extent the factor is present and relevant in the specific study (plans) by classifying it as low, midrange, or high. We advise *against* converting these levels into numbers (eg, low=1, medium=2, and high=3), as this would falsely suggest a mathematical accuracy and cardinal order and could lead to a misinterpretation.
- Step 2—*comprehensive evaluation of risks for patients:* An overview of the evaluation of the single factors leads to a comprehensive picture of the risk profile of the concrete study. At this point, the methodological question of the relationship between the individual factors arises: Is the same importance attached to each factor, or are some factors considered more important than others and therefore given greater weight when moving from the estimates of the individual factors to a more comprehensive picture? We do not consider it plausible to state a priori that some factors are more important or should count more than other factors. As a default approach to a comprehensive evaluation of risks for patients in practice, we recommend that all factors be weighted equally. In particular circumstances, practical reasoning might suggest that the estimation of a single factor as *high* still does not adequately reflect the importance of that factor in the assessment of the comprehensive evaluation of risks for patients. In this case, more weight could be given to this factor, or it could even be treated as a decisive or exclusionary criterion. This might be appropriate, for instance, if a study plans to store very large and detailed sets of personal data in an open access data repository (factor *data sharing model*).

The fact that we advise against using numbers to evaluate the individual factors (step 1) already excludes the possibility of summing up numerical values as part of a comprehensive evaluation of risks for patients and presenting the risk (step 2) in a single number. Such a numerical approach would unreasonably suggest a mathematical or empirical reliability or precision that is not justified by the framework. Instead, a comprehensive evaluation of risks for patients is based on a rough summary of the evaluations of the single factors, considering their individual weight within the evaluated study. The results of the comprehensive evaluation of risks for patients can again be presented as low, midrange, or high, depending on the distribution of the individual factors.

Use of the Tailored Risk Assessment (Steps 3-5)

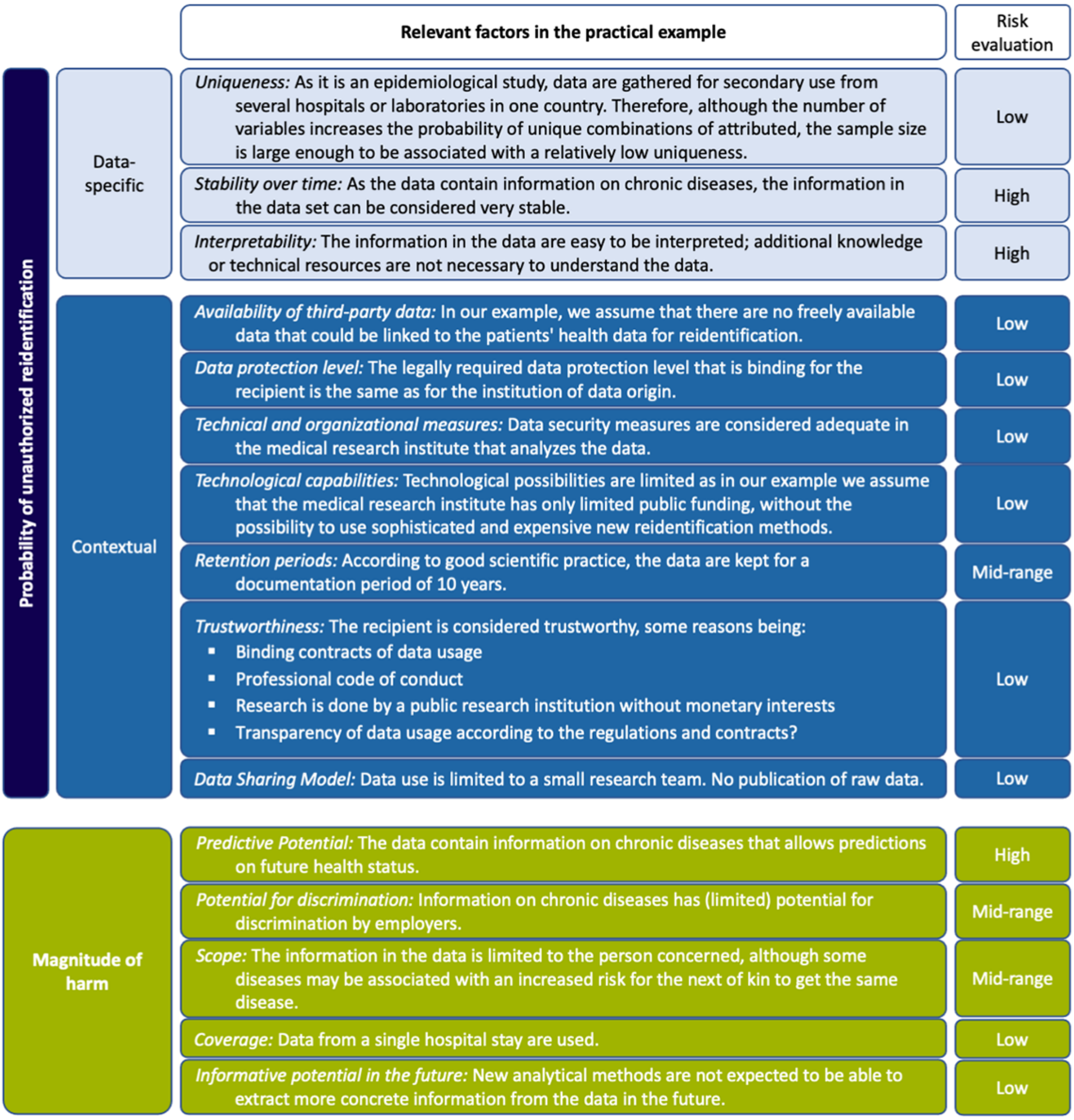
Steps 3-5 concern the use of the risk assessment framework:

- Step 3—*complete risk assessment:* It is important to be aware of the fact that the risk assessment addresses data-related risks for patients, which we deem to be the central and most important kind of risk from *SeConts*, but which are potentially still not the only kind of risk. Therefore, to gain a complete and comprehensive understanding, risks for other stakeholders (such as physicians and institutions) need to be taken into consideration.
- Step 4—*comprehensive ethical evaluation:* The complete risk assessment of a study is only one part of the comprehensive ethical evaluation. Typically, comprehensive ethical evaluation needs to include other aspects, in particular, the potential benefits of the envisaged study (to analyze and assess the risk-benefit ratio).
- Step 5—*reducing the risk profile by mitigating single factors:* The comprehensive ethical evaluation (step 4) can lead to three evaluation results of a planned study: (1) unethical and thus to be rejected, (2) ethically problematic but approvable under certain conditions, and (3) ethically sound and thus to be approved. In the case of (2), the applicant may be required to take specific measures to mitigate data-related risk to the data subjects. Tailored risk reduction measures should be chosen in light of the identification and evaluation of single risk factors (step 1) and against the backdrop of the comprehensive evaluation of risks for patients (step 2). Possible risk reduction measures may, for example, include modifying the data set to reduce the factor *uniqueness*, for example, by aggregating information (eg, age groups instead of age). In addition, special data sharing contracts can be applied to reduce the number of people who have access to the data (factor *data sharing model*).

A Practical Example of the Application of the Risk Assessment

After presenting the application and use, in the following section, we illustrate our risk assessment (steps 1 and 2) by applying it to a concrete study. Our example is a study on the epidemiology of Streptococcus pneumoniae infections [32] that we mentioned above in our list of examples for *SeConts*. According to the different areas of application (*non-interventional [observational] clinical research*, *quality control and improvement*, *public health research*) of *SeConts* we introduced above, the study can be classified as public health research. It gathers data from medical charts on “demographic characteristics, clinical syndromes, underlying conditions [eg, chronic diseases], and outcomes of illness” [32]. Figure 4 illustrates the application of the risk assessment framework for each factor (step 1) to a concrete example (For the sake of complete illustration of the risk assessment framework, we have added certain features to the study context where the study does not provide detailed information.).

Figure 4. Practical example of a study on the epidemiology of Streptococcus pneumoniae infections.



		Relevant factors in the practical example	Risk evaluation
Probability of unauthorized reidentification	Data-specific	*Uniqueness:* As it is an epidemiological study, data are gathered for secondary use from several hospitals or laboratories in one country. Therefore, although the number of variables increases the probability of unique combinations of attributed, the sample size is large enough to be associated with a relatively low uniqueness.	Low
		Stability over time: As the data contain information on chronic diseases, the information in the data set can be considered very stable.	High
		Interpretability: The information in the data are easy to be interpreted; additional knowledge or technical resources are not necessary to understand the data.	High
	Contextual	*Availability of third-party data:* In our example, we assume that there are no freely available data that could be linked to the patients' health data for reidentification.	Low
		Data protection level: The legally required data protection level that is binding for the recipient is the same as for the institution of data origin.	Low
		Technical and organizational measures: Data security measures are considered adequate in the medical research institute that analyzes the data.	Low
		Technological capabilities: Technological possibilities are limited as in our example we assume that the medical research institute has only limited public funding, without the possibility to use sophisticated and expensive new reidentification methods.	Low
		Retention periods: According to good scientific practice, the data are kept for a documentation period of 10 years.	Mid-range
		Trustworthiness: The recipient is considered trustworthy, some reasons being: ▪ Binding contracts of data usage ▪ Professional code of conduct ▪ Research is done by a public research institution without monetary interests ▪ Transparency of data usage according to the regulations and contracts?	Low
		Data Sharing Model: Data use is limited to a small research team. No publication of raw data.	Low
Magnitude of harm		*Predictive Potential:* The data contain information on chronic diseases that allows predictions on future health status.	High
		Potential for discrimination: Information on chronic diseases has (limited) potential for discrimination by employers.	Mid-range
		Scope: The information in the data is limited to the person concerned, although some diseases may be associated with an increased risk for the next of kin to get the same disease.	Mid-range
		Coverage: Data from a single hospital stay are used.	Low
		Informative potential in the future: New analytical methods are not expected to be able to extract more concrete information from the data in the future.	Low

In accordance with the application of our risk assessment approach, the results are as follows: (1) The probability of unauthorized reidentification can be considered relatively low because the majority of the relevant factors were assessed as low and (2) the magnitude of harm is in the midrange because most factors relevant to harm magnitude were assessed to be low or midrange. The comprehensive evaluation (step 2) shows that the overall risk for patients can be considered relatively low to midrange. After taking into account risks for other stakeholders (step 3), as well as the potential benefits of the study and other ethically relevant points (step 4), reviewers could request further mitigation of single risk factors (step 5). In our example, however, this is only possible to a limited extent because not all risk factors can be addressed without rendering the study itself impossible.

Limitations

Having presented our risk assessment, its application and use, and illustrating its applicability by means of a concrete example, we will address the limitations and possible criticisms of our framework. Two objections can be raised against the risk assessment approach. First, we classify a priori certain data types according to their harm potential as a basis for operationalizing the *magnitude of harm*. However, it can be argued that such an a priori classification does not take into account the information that can be inferred from the data beyond the apparent information content [72,73]: Information that is considered potentially harmful (eg, sexual orientation, religious beliefs) can be derived from information that would probably be considered harmless a priori (eg, Facebook likes) [74]. We are aware that our a priori classification of data does not consider possible inferences that can be drawn from data in the future. However, there is an important difference in whether data contain information about a certain diagnosis or whether this diagnosis can only be assumed with a certain probability. Likewise, possible inferences that can be drawn from the data can rarely be predicted a priori. Therefore, our approach is limited to identifying the types of data that have a clear potential to cause harm themselves.

A second possible criticism of the proposed approach is conceivable. Regardless of whether information is inferred indirectly from data or whether the information is contained directly in it, the same data can entail different levels of individual risk for different people. Risk assessment, such as the one we present here, cannot reflect these different levels of individual risk. Both the probability of unauthorized reidentification and the possible magnitude of harm can be very different for the same data types in different people. The x-ray of a patellar luxation may be considered as nonsensitive information for most people, especially as it does not contain any identifying information. Nevertheless, a professional soccer player might disagree, as the information in the data contains the risk of unauthorized reidentification (due to possible analog and comparable x-rays of his club) and could be potentially harmful to his career (eg, through discrimination in relation to a possible contract extension). Our approach cannot depict these individually possible risks and can only provide guidelines for the assessment of generally expected risks. The risks of individual persons with special risk profiles must be addressed using individual measures. Among other things, this calls for the establishment of a suitable information and consent procedure or an opt-out option that allows persons with an individually high risk to decide for themselves whether this risk is too high or not. Which model of informed consent may be appropriate is not the subject of this study.

Conclusions

In this paper, we addressed three desiderata of the current literature on *SeConts*. In the first step, we clarified the concept of *SeConts*. To this end, we analyzed each element of the concept and then provided a comprehensive definition of *SeConts* as *activities that solely use data produced for the sake of health care and in the context of health care to improve biomedical science or services*.

In the second step, we illustrated the scope and practical relevance of *SeConts* by providing a list of concrete types of research or learning activities that can be subsumed under the concept. These types of research or learning activities were roughly classified as either *non-interventional (observational) clinical research*, *quality control and improvement*, or *public health research*.

In the third step, we provided a framework for risk assessment for *SeConts*, focusing on the risks for patients related to informational self-determination. By operationalizing the concept of risk for application to *SeConts*, we identified factors that determine the *probability of unauthorized reidentification* as well as the *magnitude of harm* of a potential harming event implied in *SeConts*. We then discussed the application and use of our risk assessment framework and presented a practical example of a concrete study to illustrate its application.

Through our conceptual clarification of *SeConts*, we created a basis for understanding what *SeConts* means. The analysis of its scope shows that *SeConts* can realize its potential in a broad field of medical research. This illustrates the high practical relevance of *SeConts*. The risk assessment presented can be applied as an essential building block for an ethical evaluation of concrete *SeConts* conducted by research ethics committees and data use and access committees, as well as scientists, bioethicists, and funders. It can thus benefit the safe secondary use of clinical data in data-gathering, non-interventional research or learning activities.

Acknowledgments

This paper was developed within the framework of the project *Learning from Clinical Data (LinCDat)* funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation)—406103282. Project funding was awarded to ECW. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study. We would like to thank our project partners Kai Cornelius and Markus Spitz (University of Heidelberg, Germany). We also thank Nikolaus Kleindienst (Central Institute of Mental Health, Mannheim, Germany) and Adrian Thorogood (University of Luxemburg) for conceptual counseling and Vincent Lotz (National Center for Tumor Diseases, Heidelberg University Hospital, Germany) for his assistance in researching the literature and linguistic correction of the manuscript.

Authors' Contributions

MJ, CS, and ECW conceptualized the manuscript. MJ drafted the manuscript, and CS supervised the preparation. AK and KM have revised the manuscript.

Conflicts of Interest

None declared.

References

1. Warren-Gash C. Linking and sharing routine health data for research in England. Cambridge: PHG Foundation; 2017:1-25.

2. Institute of Medicine, Committee on the Learning Health Care System in America, Smith M, Saunders R, Stuckhardt L, McGinnis J. Best care at lower cost: the path to continuously learning health care in America. Washington, D.C: The National Academies Press; 2013.
3. Schilsky RL, Michels DL, Kearbey AH, Yu PP, Hudis CA. Building a rapid learning health care system for oncology: the regulatory framework of CancerLinQ. J Clin Oncol 2014 Aug 01;32(22):2373-2379. [doi: 10.1200/jco.2014.56.2124]
4. Martin-Sanchez FJ, Aguiar-Pulido V, Lopez-Campos GH, Peek N, Sacchi L. Secondary use and analysis of big data collected for patient care. Yearb Med Inform 2017 Aug 19;26(01):28-37. [doi: 10.1055/s-0037-1606529]
5. Brakewood B, Poldrack RA. The ethics of secondary data analysis: considering the application of Belmont principles to the sharing of neuroimaging data. Neuroimage 2013 Nov 15;82:671-676. [doi: 10.1016/j.neuroimage.2013.02.040] [Medline: 23466937]
6. Lucero RJ, Kearney J, Cortes Y, Arcia A, Appelbaum P, Fernández RL, et al. Benefits and risks in secondary use of digitized clinical data: views of community members living in a predominantly ethnic minority urban neighborhood. AJOB Empir Bioeth 2015 Sep 11;6(2):12-22 [FREE Full text] [doi: 10.1080/23294515.2014.949906] [Medline: 26101782]
7. Elger BS, Iavindrasana J, Lo Iacono L, Müller H, Roduit N, Summers P, et al. Strategies for health data exchange for secondary, cross-institutional clinical research. Comput Methods Programs Biomed 2010 Sep;99(3):230-251. [doi: 10.1016/j.cmpb.2009.12.001] [Medline: 20089327]
8. Lee LM. Ethics and subsequent use of electronic health record data. J Biomed Inform 2017 Jul;71:143-146 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.022] [Medline: 28578074]
9. Aven T. On how to define, understand and describe risk. Reliab Eng Syst Safe 2010 Jun;95(6):623-631. [doi: 10.1016/j.ress.2010.01.011]
10. Alaszewski A, Harrison L, Manthorpe J. Risk, Health and Welfare: Policies, Strategies and Practice. Milton Keynes: Open University Press; 1998:1-176.
11. Nida-Rümelin J. Ethik des Risikos. Die Bereichsethiken und ihre theoretische Fundierung. In: Nida-Rümelin J, editor. Angewandte Ethik. Stuttgart: Alfred Kröner Verlag; 2005.
12. Verizon. Verizon: 2019 Data Breach Investigations Report. Computer Fraud & Security 2019 Jun;2019(6):4. [doi: 10.1016/S1361-3723(19)30060-0]
13. Floyd T, Grieco M, Reid E. Mining hospital data breach records: cyber threats to U.S. Hospitals. In: Proceedings of the IEEE Conference on Intelligence and Security Informatics (ISI). 2016 Presented at: IEEE Conference on Intelligence and Security Informatics (ISI); Sept. 28-30, 2016; Tucson, AZ, USA. [doi: 10.1109/isi.2016.7745441]
14. Liu V, Musen MA, Chou T. Data breaches of protected health information in the United States. J Am Med Assoc 2015 Apr 14;313(14):1471-1473 [FREE Full text] [doi: 10.1001/jama.2015.2252] [Medline: 25871675]
15. Datenschutz-Folgenabschätzung. Methodik und Fallstudie. Der Bayerische Landesbeauftragte für den Datenschutz. 2019. URL: https://www.datenschutz-bayern.de/technik/orient/oh_dsfa_beispiel.pdf [accessed 2021-06-01]
16. Bitkom - AK Datenschutz. Bitcom e.V. 2017. URL: https://www.bitkom.org [accessed 2021-05-26]
17. Datenschutzgruppe nach Artikel 29. Leitlinien zur Datenschutz-Folgenabschätzung (DSFA) und Beantwortung der Frage, ob eine Verarbeitung im Sinne der Verordnung 2016/679 "wahrscheinlich ein hohes Risiko mit sich bringt?". 2017. URL: https://www.datenschutz-bayern.de/technik/orient/wp248.pdf [accessed 2021-05-26]
18. Rid A, Emanuel EJ, Wendler D. Evaluating the risks of clinical research. J Am Med Assoc 2010 Oct 06;304(13):1472-1479. [doi: 10.1001/jama.2010.1414] [Medline: 20924013]
19. Rid A, Wendler D. A framework for risk-benefit evaluations in biomedical research. Kennedy Inst Ethics J 2011 Jun;21(2):141-179. [doi: 10.1353/ken.2011.0007] [Medline: 21696094]
20. Dyke SO, Dove ES, Knoppers BM. Sharing health-related data: a privacy test? NPJ Genom Med 2016 Aug 17;1(1):160241-160246 [FREE Full text] [doi: 10.1038/npjgenmed.2016.24] [Medline: 27990299]
21. Bundesministerium für Bildung und Forschung. Medizininformatik Initiative. URL: https://www.medizininformatik-initiative.de [accessed 2021-05-26]
22. Rumbold JM, Pierscionek BK. What are data? A categorization of the data sensitivity spectrum. Big Data Res 2018 Jul;12:49-59. [doi: 10.1016/j.bdr.2017.11.001]
23. Köngeter A, Jungkunz M, Winkler EC, Schickhardt C, Mehlis K. Sekundärnutzung klinischer Daten aus der Patientenversorgung für Forschungszwecke - Eine qualitative Interviewstudie zu Nutzen- und Risikopotenzialen aus Sicht von Expertinnen und Experten für den deutschen Forschungskontext. In: Richter G, Loh W, Buyx A, Kielmansegg S, editors. Datenreiche Medizin und das Problem der Einwilligung. Ethische, rechtliche und sozialwissenschaftliche Perspektiven. Heidelberg: Springer; 2021.
24. Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Expert Panel. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc 2007;14(1):1-9 [FREE Full text] [doi: 10.1197/jamia.M2273] [Medline: 17077452]
25. Robertson AR, Nurmatov U, Sood HS, Cresswell K, Smith P, Sheikh A. A systematic scoping review of the domains and innovations in secondary uses of digitised health-related data. J Innov Health Inform 2016 Nov 10;23(3):611-619 [FREE Full text] [doi: 10.14236/jhi.v23i3.841] [Medline: 28059695]

26. American Medical Informatics Association. Secondary uses and re-uses of healthcare data: taxonomy for policy formulation and planning. 2007. URL: https://www.amia.org/sites/amia.org/files/2007-Policy-Meeting-amia-taxonomy-Secondary-Data-Use-Version.pdf [accessed 2021-05-26]
27. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: a scoping review. PLoS One 2018 Oct 11;13(10):e0204937 [FREE Full text] [doi: 10.1371/journal.pone.0204937] [Medline: 30308031]
28. Brook E, Rosman D, Holman C. Public good through data linkage: measuring research outputs from the Western Australian Data Linkage System. Aust N Z J Public Health 2008 Feb;32(1):19-23. [doi: 10.1111/j.1753-6405.2008.00160.x] [Medline: 18290908]
29. Evans RS, Burke JP, Classen DC, Gardner RM, Menlove RL, Goodrich KM, et al. Computerized identification of patients at high risk for hospital-acquired infection. Am J Infect Control 1992 Feb;20(1):4-10. [doi: 10.1016/s0196-6553(05)80117-8]
30. Pittet D, Safran E, Harbarth S, Borst F, Copin P, Rohner P, et al. Automatic alerts for methicillin-resistant surveillance and control: role of a hospital information system. Infect Control Hosp Epidemiol 2015 Jan 02;17(8):496-502. [doi: 10.1017/s0195941700004641]
31. Smith G, Hippisley-Cox J, Harcourt S, Heaps M, Painter M, Porter A, et al. Developing a national primary care-based early warning system for health protection--a surveillance tool for the future? Analysis of routinely collected data. J Public Health (Oxf) 2007 Mar;29(1):75-82. [doi: 10.1093/pubmed/fdl078] [Medline: 17158478]
32. Robinson KA, Baughman W, Rothrock G, Barrett NL, Pass M, Lexau C, Active Bacterial Core Surveillance (ABCs)/Emerging Infections Program Network. Epidemiology of invasive Streptococcus pneumoniae infections in the United States, 1995-1998: opportunities for prevention in the conjugate vaccine era. J Am Med Assoc 2001 Apr 04;285(13):1729-1735. [doi: 10.1001/jama.285.13.1729] [Medline: 11277827]
33. Been JV, Mackay DF, Millett C, Pell JP, van Schayck OC, Sheikh A. Impact of smoke-free legislation on perinatal and infant mortality: a national quasi-experimental study. Sci Rep 2015 Aug 13;5(1):13020 [FREE Full text] [doi: 10.1038/srep13020] [Medline: 26268789]
34. Baker DW, Persell SD, Thompson JA, Soman NS, Burgner KM, Liss D, et al. Automated review of electronic health records to assess quality of care for outpatients with heart failure. Ann Intern Med 2007 Feb 20;146(4):270-277. [doi: 10.7326/0003-4819-146-4-200702200-00006] [Medline: 17310051]
35. Hay JW, Hay AR. Inflammatory bowel disease: costs-of-illness. J Clin Gastroenterol 1992 Jun;14(4):309-317. [doi: 10.1097/00004836-199206000-00009] [Medline: 1607607]
36. Stattin P, Holmberg E, Johansson J, Holmberg L, Adolfsson J, Hugosson J, National Prostate Cancer Register (NPCR) of Sweden. Outcomes in localized prostate cancer: National Prostate Cancer Register of Sweden follow-up study. J Natl Cancer Inst 2010 Jul 07;102(13):950-958 [FREE Full text] [doi: 10.1093/jnci/djq154] [Medline: 20562373]
37. Weiner M, Xie D, Tannen R. Clinical trials in silico: rigorous assessment of treatment effect using electronic health records. AMIA Annu Symp Proc 2008 Nov 06:1172. [Medline: 18999175]
38. Bronsert MR, Henderson WG, Valuck R, Hosokawa P, Hammermeister K. Comparative effectiveness of antihypertensive therapeutic classes and treatment strategies in the initiation of therapy in primary care patients: a Distributed Ambulatory Research in Therapeutics Network (DARTNet) study. J Am Board Fam Med 2013;26(5):529-538 [FREE Full text] [doi: 10.3122/jabfm.2013.05.130048] [Medline: 24004705]
39. Stürmer T, Funk MJ, Poole C, Brookhart MA. Nonexperimental comparative effectiveness research using linked healthcare databases. Epidemiology 2011 May;22(3):298-301 [FREE Full text] [doi: 10.1097/EDE.0b013e318212640c] [Medline: 21464649]
40. Oshikoya KA, Wharton GT, Avant D, Van Driest SL, Fenn NE, Lardieri A, et al. Serious adverse events associated with off-label use of azithromycin or fentanyl in children in intensive care units: a retrospective chart review. Paediatr Drugs 2019 Feb;21(1):47-58 [FREE Full text] [doi: 10.1007/s40272-018-0318-9] [Medline: 30484072]
41. Kuter DJ, Mehta A, Hollak CE, Giraldo P, Hughes D, Belmatoug N, et al. Miglustat therapy in type 1 Gaucher disease: clinical and safety outcomes in a multicenter retrospective cohort study. Blood Cells Mol Dis 2013 Aug;51(2):116-124. [doi: 10.1016/j.bcmd.2013.04.005] [Medline: 23683771]
42. Mathews JD, Forsythe AV, Brady Z, Butler MW, Goergen SK, Byrnes GB, et al. Cancer risk in 680,000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. Br Med J 2013 May 21;346:f2360 [FREE Full text] [doi: 10.1136/bmj.f2360] [Medline: 23694687]
43. Conway PH, Cnaan A, Zaoutis T, Henry BV, Grundmeier RW, Keren R. Recurrent urinary tract infections in children: risk factors and association with prophylactic antimicrobials. J Am Med Assoc 2007 Jul 11;298(2):179-186. [doi: 10.1001/jama.298.2.179] [Medline: 17622599]
44. Yeniterzi R, Aberdeen J, Bayer S, Wellner B, Hirschman L, Malin B. Effects of personal identifier resynthesis on clinical text de-identification. J Am Med Inform Assoc 2010;17(2):159-168 [FREE Full text] [doi: 10.1136/jamia.2009.002212] [Medline: 20190058]
45. Köpcke F, Kraus S, Scholler A, Nau C, Schüttler J, Prokosch H, et al. Secondary use of routinely collected patient data in a clinical trial: an evaluation of the effects on patient recruitment and data acquisition. Int J Med Inform 2013 Mar;82(3):185-192. [doi: 10.1016/j.ijmedinf.2012.11.008] [Medline: 23266063]

46. Schickhardt C, Fleischer H, Winkler EC. Do patients and research subjects have a right to receive their genomic raw data? An ethical and legal analysis. BMC Med Ethics 2020 Jan 16;21(1):7 [FREE Full text] [doi: 10.1186/s12910-020-0446-y] [Medline: 31948449]
47. Rössler B. Der Wert des Privaten. Frankfurt am Main: Suhrkamp; 2001.
48. Fried C. Privacy. Yale Law J 1968 Jan;77(3):475-493. [doi: 10.2307/794941]
49. Solove DJ. Understanding Privacy. Cambridge, MA: Harvard University Press; 2008.
50. de Bruin B. The liberal value of privacy. Law Philos 2010 Apr 10;29(5):505-534. [doi: 10.1007/s10982-010-9067-9]
51. Rath C. Karlsruhe und der Einschüchterungseffekt? Praxis und Nutzen einer Argumentationsfigur des Bundesverfassungsgerichts. In: Verfassungsrecht und gesellschaftliche Realität. Baden-Baden, Germany: Nomos; 2009:65-80.
52. Kaplan B. How Should Health Data Be Used? Camb Q Healthc Ethics 2016 Mar 09;25(2):312-329. [doi: 10.1017/s0963180115000614]
53. Agaku IT, Adisa AO, Ayo-Yusuf OA, Connolly GN. Concern about security and privacy, and perceived control over collection and use of health information are related to withholding of health information from healthcare providers. J Am Med Inform Assoc 2014;21(2):374-378 [FREE Full text] [doi: 10.1136/amiajnl-2013-002079] [Medline: 23975624]
54. Benitez K, Malin B. Evaluating re-identification risks with respect to the HIPAA privacy rule. J Am Med Inform Assoc 2010;17(2):169-177 [FREE Full text] [doi: 10.1136/jamia.2009.000026] [Medline: 20190059]
55. Laurie G, Jones K, Stevens L, Dobbs C. A review of evidence relating to harm resulting from uses of health and biomedical data. Nuffield Council on Bioethics. 2014 Feb 15. URL: https://www.research.ed.ac.uk/en/publications/a-review-of-evidence-relating-to-harm-resulting-from-uses-of-heal [accessed 2021-05-31]
56. Bender S, Brand R, Bacher J. Re-identifying register data by survey data: an empirical study. Stat J Unit Nation 2001 Dec 28;18(4):373-381. [doi: 10.3233/sju-2001-18412]
57. Mokken R, Kooiman P, Pannekoek J, Willenborg L. Disclosure risks for microdata. Stat Neerl 1992 Mar;46(1):49-67. [doi: 10.1111/j.1467-9574.1992.tb01326.x]
58. Dankar FK, El Emam K, Neisa A, Roffey T. Estimating the re-identification risk of clinical data sets. BMC Med Inform Decis Mak 2012 Jul 09;12:66 [FREE Full text] [doi: 10.1186/1472-6947-12-66] [Medline: 22776564]
59. Emam KE, Dankar FK, Vaillancourt R, Roffey T, Lysyk M. Evaluating the risk of re-identification of patients from hospital prescription records. Can J Hosp Pharm 2009 Jul;62(4):307-319 [FREE Full text] [doi: 10.4212/cjhp.v62i4.812] [Medline: 22478909]
60. Sweeney L. K-anonymity: a model for protecting privacy. Int J Unc Fuzz Knowl Based Syst 2012 May 02;10(05):557-570. [doi: 10.1142/S0218488502001648]
61. Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M. L-diversity: privacy beyond k-anonymity. In: Proceedings of the 22nd International Conference on Data Engineering (ICDE'06). 2006 Presented at: 22nd International Conference on Data Engineering (ICDE'06); April 3-7, 2006; Atlanta, GA, USA. [doi: 10.1109/ICDE.2006.1]
62. Li N, Li T, Venkatasubramanian S. T-closeness: privacy beyond k-anonymity and l-diversity. In: Proceedings of the IEEE 23rd International Conference on Data Engineering. 2007 Presented at: 2007 IEEE 23rd International Conference on Data Engineering; April 15-20, 2007; Istanbul, Turkey. [doi: 10.1109/icde.2007.367856]
63. Malin B, Loukides G, Benitez K, Clayton EW. Identifiability in biobanks: models, measures, and mitigation strategies. Hum Genet 2011 Sep;130(3):383-392 [FREE Full text] [doi: 10.1007/s00439-011-1042-5] [Medline: 21739176]
64. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science 2013 Jan 18;339(6117):321-324 [FREE Full text] [doi: 10.1126/science.1229566] [Medline: 23329047]
65. Ellenbogen P, Narayanan A. Identification of anonymous DNA using genealogical triangulation. bioRxiv 2019:531269. [doi: 10.1101/531269]
66. Golle P. Revisiting the uniqueness of simple demographics in the US population. In: Proceedings of the 5th ACM Workshop on Privacy in Electronic Society. 2006 Presented at: CCS06: 13th ACM Conference on Computer and Communications Security 2006; 30 October, 2006; Alexandria Virginia USA p. 77-80. [doi: 10.1145/1179601.1179615]
67. Arbuckle L, Ritchie F. The five safes of risk-based anonymization. IEEE Secur Privacy 2019 Sep;17(5):84-89. [doi: 10.1109/msec.2019.2929282]
68. Prasser F, Kohlmayer F, Kuhn K. The importance of context: risk-based de-identification of biomedical data. Methods Inf Med 2018 Jan 08;55(04):347-355. [doi: 10.3414/me16-01-0012]
69. Naveed M, Ayday E, Clayton EW, Fellay J, Gunter CA, Hubaux J, et al. Privacy in the genomic era. ACM Comput Surv 2015 Sep;48(1):6 [FREE Full text] [doi: 10.1145/2767007] [Medline: 26640318]
70. Schuol S, Winkler EC. Umgang mit genomischen Daten in der klinischen Forschung aus ethischer Sicht. Forum 2016 Sep 29;31(5):402-405. [doi: 10.1007/s12312-016-0124-7]
71. EURAT. Eckpunkte für eine Heidelberger Praxis der Ganzgenomsequenzierung. Heidelberg: Marsilius-Kolleg der Universität Heidelberg; 2013.
72. Hagendorff T. From privacy to anti-discrimination in times of machine learning. Ethics Inf Technol 2019 Aug 5;21(4):331-343. [doi: 10.1007/s10676-019-09510-5]
73. Wachter S, Mittelstadt B. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Colum Bus L Rev 2019;2:494-620. [doi: 10.31228/osf.io/mu2kf]

74. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci U S A 2013 Apr 09;110(15):5802-5805 [FREE Full text] [doi: 10.1073/pnas.1218772110] [Medline: 23479631]

Abbreviations

DFG: Deutsche Forschungsgemeinschaft
EHR: electronic health record
GDPR: General Data Protection Regulation
LHCS: learning health care system
LinCDat: Learning from Clinical Data
SeConts: secondary use of clinical data in data-gathering, non-interventional research or learning activities

Edited by R Kukafka; submitted 21.12.20; peer-reviewed by P Swire, L Heryawan; comments to author 08.02.21; revised version received 10.03.21; accepted 06.05.21; published 08.06.21.

Please cite as:
Jungkunz M, Köngeter A, Mehlis K, Winkler EC, Schickhardt C
Secondary Use of Clinical Data in Data-Gathering, Non-Interventional Research or Learning Activities: Definition, Types, and a Framework for Risk Assessment
J Med Internet Res 2021;23(6):e26631
URL: https://www.jmir.org/2021/6/e26631
doi:10.2196/26631
PMID:34100760

Original Paper

Consumer Views on Health Applications of Consumer Digital Data and Health Privacy Among US Adults: Qualitative Interview Study

David Grande[1,2], MD, MPA; Xochitl Luna Marti[1], MPH; Raina M Merchant[3], MD, MSHP; David A Asch[1], MD, MBA; Abby Dolan[3], MPH; Meghana Sharma[1], MPH; Carolyn C Cannuscio[4], SCD

[1]Division of General Internal Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA, United States

[4]Department of Family Medicine and Community Health, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
David Grande, MD, MPA
Division of General Internal Medicine
University of Pennsylvania
3641 Locust Walk
CPC 407
Philadelphia, PA, 19081
United States
Phone: 1 2155733804
Fax: 1 2158980229
Email: dgrande@wharton.upenn.edu

Abstract

Background: In 2020, the number of internet users surpassed 4.6 billion. Individuals who create and share digital data can leave a trail of information about their habits and preferences that collectively generate a digital footprint. Studies have shown that digital footprints can reveal important information regarding an individual's health status, ranging from diet and exercise to depression. Uses of digital applications have accelerated during the COVID-19 pandemic where public health organizations have utilized technology to reduce the burden of transmission, ultimately leading to policy discussions about digital health privacy. Though US consumers report feeling concerned about the way their personal data is used, they continue to use digital technologies.

Objective: This study aimed to understand the extent to which consumers recognize possible health applications of their digital data and identify their most salient concerns around digital health privacy.

Methods: We conducted semistructured interviews with a diverse national sample of US adults from November 2018 to January 2019. Participants were recruited from the Ipsos KnowledgePanel, a nationally representative panel. Participants were asked to reflect on their own use of digital technology, rate various sources of digital information, and consider several hypothetical scenarios with varying sources and health-related applications of personal digital information.

Results: The final cohort included a diverse national sample of 45 US consumers. Participants were generally unaware what consumer digital data might reveal about their health. They also revealed limited knowledge of current data collection and aggregation practices. When responding to specific scenarios with health-related applications of data, they had difficulty weighing the benefits and harms but expressed a desire for privacy protection. They saw benefits in using digital data to improve health, but wanted limits to health programs' use of consumer digital data.

Conclusions: Current privacy restrictions on health-related data are premised on the notion that these data are derived only from medical encounters. Given that an increasing amount of health-related data is derived from digital footprints in consumer settings, our findings suggest the need for greater transparency of data collection and uses, and broader health privacy protections.

(J Med Internet Res 2021;23(6):e29395) doi:10.2196/29395

KEYWORDS

health privacy; digital health privacy; privacy law; health law; digital epidemiology

Introduction

In 2020, internet users spent 1.25 billion years online. The average internet user spends 6 hours and 43 minutes online each day [1], leaving a trail of information about his/her habits and preferences that collectively generates a digital footprint.

A consumer's digital footprint can reveal health-related behaviors, such as diet and physical activity, that can predict health [2]. For example, data from social media sites, including Twitter and Facebook, can be used to screen for signs of depression, suicidal ideation, and sleep disorders based on user activity and language patterns [3-6]. These applications of data toward health purposes erase past distinctions between health and nonhealth data [7,8].

The COVID-19 pandemic has accelerated these applications. Smartphone location services and cashless transactions have been used to identify and track COVID-19–positive individuals [9]. Social media posts have been used to track reports of physical symptoms of COVID-19 and mental health sequelae of the pandemic [10,11]. Apple and Google have updated their smartphone operating systems to enable tracking of human-to-human interactions to enable digital contact tracing [12]. These public health applications of consumer digital data have triggered global debates and urgent policy discussions about digital health privacy [13].

A majority of Americans simultaneously reported feeling concerned about how their personal data are being used and yet continue to use digital technologies. Many believe it is impossible to go through daily life without having personal data collected by companies or the government, with 80% feeling they lack control over what data are collected and 63% admitting having little to no understanding about data privacy laws and regulations [14].

We interviewed a diverse sample of American consumers to evaluate their awareness of the health applications of their digital data and identify their privacy views and concerns.

Methods

Participants

Interviews were conducted between November 1, 2018, and January 31, 2019. Participants were identified from the web-enabled Ipsos KnowledgePanel, a probability-based panel designed to be representative of the US population [15]. For this study, Ipsos provided the contact information of 200 participants drawn from the nationally representative panel with even distribution for various demographic variables (gender, race, age, household income, and geographic region) to ensure diversity in the study sample. From these, team members (XLM and AL) completed 45 interviews, ensuring equal proportions of participants were included from each category so that the sample of 45 reflected the larger sample of 200. Participants who completed the interview were paid US $50.

This study protocol was reviewed by the Institutional Review Board of the University of Pennsylvania and declared exempt. Verbal consent was obtained from participants prior to their interview. All results are reported following the Standards for Reporting Qualitative Research (SRQR) reporting guidelines.

Design

Telephone interviews were conducted using a semistructured qualitative interview guide (Multimedia Appendix 1). This guide was informed by a consequential ethics framework, in which the presence or absence of a substantial risk of harm from a loss of privacy determines the need for protections [16]. The guide asked consumers to reflect on their own use of digital technology, evaluate sources of digital information, and consider several hypothetical scenarios with varying uses of digital information (scenarios shown in Table 1). While the scenarios were hypothetical, they were developed based on plausible use cases in today's landscape of consumer digital data. The interviews lasted 30 to 45 minutes, and were conducted by research coordinators (AL and XLM) trained in qualitative interviewing methods. The audio recordings were sent to a professional transcription service (ADA Transcription) where they were deidentified. Transcripts were uploaded to NVivo Version 12 for data analysis [17].

Table 1. Scenarios.

No.	Text	Source	User	Use
1	*Your health insurance company is trying to find ways to keep people healthier and save money. They have found that consumers that buy certain kinds of food are more likely to develop diabetes. The insurance company is planning a program where they will access the grocery shopping records of their patients from grocery stores. The health insurance company will use this information to find out who is at high risk of developing diabetes, then send those people tips and advice on how they can prevent diabetes by making changes to the food they buy.*	Grocery shopping records	Health insurance company	Prevent diabetes
2	*Your doctor's office is trying to find ways to prevent people from getting sick and needing to go to the hospital. They have found that patients that search on the internet for certain symptoms are more likely to get sick and need to go to an emergency room. Your doctor's office is planning a new program where they will access internet searches of their patients and contact patients that search for certain symptoms to try to start treatment sooner.*	Internet search history	Doctor's office	Diagnose disease earlier
3	*University researchers are trying to find ways to prevent cancer. Researchers at a nearby university hospital are starting a research study where they will track patients over time to try to determine causes of cancer. In addition to using medical records, the research team will use location information from patient's smartphones so they can study how the places where people spend most of their time impact their risk of getting cancer. The researchers want to use this knowledge to help develop public health strategies in the future that could reduce the number of people with cancer.*	Electronic health record and location data from a smartphone	University researchers	Research to identify risk factors for cancer
4	*DigiHealth is a company selling a new smartphone app that can automatically collect and store information on places users visit and the food they eat so that it can give advice on ways to lower their risk of obesity. The app tracks where users go using location services on their smartphone and tracks what they eat by having them upload a picture of their meals. DigiHealth is able to offer the app for free because it shares user information with advertisers so they can send out grocery coupons.*	Location data and photos from a smartphone	Technology company	Send grocery coupons for healthier foods to prevent obesity

Analysis

The study team developed a codebook through line-by-line iterative reading and notation of transcripts, which produced 12 key codes [18]. Some codes were used a priori in order to compare data from consumers to the data from similar interviews conducted with experts [7]. Two research coordinators used the codebook to complete coding (AL and XLM). In order to establish agreement, five interview transcripts were double coded. Interrater reliability was high, with overall agreement (97.5%) and Cohen kappa of 0.7. After agreement was established, the remaining transcripts were coded individually. Interviews were coded sequentially in the order they were conducted. A memo was written, and it thematically summarized the content of each code. These memos were reviewed collectively by the study team in order to identify salient findings. The study team found that they were no longer learning new information in the process of coding new transcripts. Therefore, through coding and analysis meetings, key study personnel reached a consensus that thematic saturation was reached. The results from these interviews are reported along with supporting quotes to highlight relevant findings.

Results

Participants

The sampling methods produced a final cohort of 45 consumers with diversity in gender, race, age, household income, and geographical distribution (Table 2).

Table 2. Demographics of the participants (N=45).

Characteristic	Value, n (%)
Gender	
Male	24 (53%)
Female	21 (47%)
Race	
Minority (non-White)	22 (49%)
White/Caucasian	23 (51%)
Age (years)	
18-30	14 (31%)
31-45	10 (22%)
46-60	11 (24%)
>60	10 (22%)
Household income (US$)	
≤24,999	10 (22%)
25,000-49,999	11 (24%)
50,000-99,999	13 (29%)
≥100,000	11 (24%)
Region	
Midwest	12 (27%)
Northeast	10 (22%)
South	13 (29%)
West	10 (22%)
Education	
Less than high school	3 (6%)
High school	12 (27%)
Some college	13 (29%)
Bachelor’s degree or higher	17 (38%)

Part 1: Views on Digital Privacy and Health

The first portion of participant interviews focused on consumers’ understanding of current data use as well as their views on digital health privacy. Key themes are described below, and illustrative quotes are presented in Table 3.

Table 3. Key themes from consumer interviews.

Theme	Illustrative quote
Lack of consumer understanding	*I don't really have a great knowledge of what they gather from me. I know when I signed up and you gotta go through the cursory that says do you agree to all this, there's just pages of stuff and you just say, like everything else, you just say yeah, I agree.* *I don't think Google collects any information about me when I use it.* *My health? I mean, they wouldn't think I have any health issue...I don't make any payments in regards to my health .* *I don't think they could learn nothing. You don't put your health information on there. Like what would they know from Instagram?* *I don't necessarily know what they would gain about my health overall other than sometimes when I'm sick, I'll research my symptoms online or something like that.*
Benefits	*Well, I think the biggest benefit that I can think of –the reason that I use Reddit is because it's beneficial to feel like I belong to a handful of communities. And so, I think I deepen my interest –the things that I'm interested in feel more meaningful when I read what other people are talking about.* *I think it would help them –if they were gathering that information, they would use it for marketing and potential sales, how they go about structuring better ways to advertise to persons in organizations.* *Well, benefits –I get to find out things that my friends are doing that they want me to know, because otherwise they wouldn't post it. But it's also a double-edged sword in that –basically, that –that those –whatever they're talking about –whatever –however I'm reacting to it, is more information for advertisers to target me with.*
Harms	*Well, I think that information should just be between me and my doctor. I wouldn't want any advertisers to be able to learn that information, and I wouldn't want anything that I wouldn't share with my family or friends to come out in the public regarding anything that was physically going on with me.* *Well, I think the –some of the worry is that the Fitbit–I don't have one or use one, but –so the worry is that it might provide inaccurate data that might be used against me. For example, if I applied for life insurance, I would worry that if the insurance company got a hold of Fitbit records out of context that they could –I could wind up with a higher premium or something like that.* *That's just sensitive private information. Whether I have a cold or whether I have a more serious ailment disease, I think that's private. Unless I wanna share it with people, that shouldn't be disclosed to anyone without my expressed consent.*
Consumer attitudes	*I really don't have nothing to hide, but that's pretty personal to a lot of people, and I really wouldn't like it. So I would just –I mean, you've got to have a little bit of freedom here. That's what this country's about.* *I think the worst ones would be my health bills. If that was not kept private, that would be really problematic. If you think about cell phone bills, electricity bills, internet bills, I don't think I would have as much as an issue.*

Lack of Consumer Understanding of Digital Data Practices

Few participants could express a basic understanding of digital data collection practices, including how information is routinely tracked, stored, and subsequently used by third parties. Some were unaware that by downloading and using an app (eg, Facebook), they had consented to data collection. Participants who were aware that their data were being collected generally assumed the data were being used only by the entity collecting it and were not shared with third parties. Almost none of the participants named ways in which data are aggregated across sources and over time, a common practice by marketers to build profiles of individual consumers.

Participants generally did not recognize the connections between their digital data and inferences about their health. They were not able to conceptualize health information outside of a traditional definition of health care data. They were unaware that predictions or inferences about their health status could be drawn from sources like internet browsing information or social media posts.

Perceived Benefits and Harms of Digital Data Practices

Participants noted benefits of digital data collection to consumers and companies. They described the benefits of advertising highly tailored to their interests. They discussed convenience derived from data tracking. They discussed health benefits when prompted by specific health-related scenarios, but rarely mentioned them during other parts of the interview despite the interviewers introducing the study focus as related to health.

Indeed, most privacy concerns were broad rather than specific to health. Participants worried that their information could affect their finances, particularly that insurance coverage could be lost or premiums raised because of health information being made available to their insurer. Some participants feared losing employment opportunities if their information, especially health information, was available to an employer.

For some participants, the main harm was the intrusiveness of the digital data collection practices of the government, commercial companies like health insurers, and even health care providers, with access to information that some felt should remain private.

Participants split into two points of view when weighing tradeoffs between the benefits and risks of digital data collection and use. Some felt that it was a choice to give up privacy for convenience or other benefits. Several of these participants discussed tradeoffs in transactional terms. The use of discount cards, apps, and other programs naturally involves relinquishing some privacy, which is sometimes a worthwhile choice. Others described data tracking and sharing as inescapable. These participants did not consider these privacy risks acceptable but struggled to articulate how to avoid them.

Attitudes Regarding Digital Privacy

Despite a lack of knowledge on data privacy issues, participants strongly endorsed the significance of data privacy, based partly on a right to privacy and on concerns about harms.

When asked about privacy preferences for different types of data (health information, internet searches, and financial information), participants consistently expressed the greatest concern if their health information was not kept private. However, participants did not understand that data generated outside of health care could also be used to make inferences about their health.

Part 2: Health-Related Applications of Consumer Digital Data

To make health applications of digital data more salient to participants, we prompted them with a series of hypothetical scenarios (Table 1) representing various sources, uses, and users of digital information for health purposes. For each scenario, participants were asked to provide their overall impression, describe what they liked and disliked about it, identify limits or protections they desired, and share whether they would want to participate.

Although participants' responses to each scenario varied, certain themes were consistent. Participants appreciated the health benefits that could be achieved through the programs, including interventions to improve individual or population health, research to make advances in health care, and efforts to reduce health care costs. However, participants often expressed concern that many uses seemed too invasive and that their data might be used for other purposes beyond what was intended. They often desired protections so that data could be used only for the originally intended purpose, data could be kept private and secure, and participation could be voluntary and require consent. Table 4 provides illustrative quotes for the benefits, risks, and protections discussed by participants for each scenario.

Table 4. Consumer perspectives on digital data use scenarios (illustrative quotes).

Scenario[a] (% favorable)	Rated favorable (N=45), n (%)	Benefit quotes	Risk quotes	Desired protection quotes
A health insurance company accesses data on consumers' grocery store purchases to identify individuals at high risk for developing diabetes and sends them healthy eating tips.	20 (44%)	*I think it could give some valuable information. And especially if someone is unaware that their purchases are causing problems.* *The idea that someone is going to help me because I might not have that information that this product might not be good for my health, that is wonderful.*	*How much do you want them to stick their nose into what you're doing? … Now they want to tell you what to eat? They could easily raise the premium on it. I know they would.* *That's just my personal choice if that's what I wanna purchase and I don't need my health insurance charging me...*	*It's just I don't want them using it for different purposes, other than what … they were gonna use it for.* *I don't think they should be able to use your –that information against you as far as, again, raising premiums or copays or deciding what's a covered cost and what's not a covered cost based on my lifestyle as far as what I eat.*
A doctor's office monitors internet search data of their patients and contacts individuals who search for symptoms of illness in order to provide an early diagnosis.	12 (27%)	*…The doctor can be like hey, I seen you looked this up about measles and now all of a sudden you think you have measles –it's all in your head.* *They can say, oh, I see that you're looking for information on a child that's throwing up chronically. Let's get him in here and check him out...*	*They shouldn't have access to what I'm searching for online because I am a Web M.D. queen…I will Google the symptoms and see if it's something else or what I can do before I go to the doctor.* *It's just an invasion of privacy. Again, if you are volunteering that information, then you can volunteer whatever information you want.*	*It would have to be on an opt-in basis so that people are made aware of what's gonna happen and they're choosing to let that happen* *That should definitely be protected, because that's sensitive.*
University researchers access the smartphone location and medical records of patients to study risk factors for cancer.	35 (78%)	*I mean, that idea is something worth being monitored for because it's for a good cause. I mean, it can potentially save hundreds of lives, and it can also save insurance companies.*	*Just I don't like anyone knowing my whereabouts all hours of the day and all night, especially – and it's not even just anyone. It's – a stranger…* *Personally, no, because I don't want them thinking because of where I work at that I'm buying tobacco all the time because I work at a store that sells tobacco and alcohol.*	*If there was a situation where I thought that someplace I went would give the wrong idea out of context, then just turn off GPS.* *Just knowing that they are not allowed to sell my information for anything else other than using it specifically for this research.*
A technology company creates an app that tracks users' locations and asks them to upload photos of their meals with the intention of encouraging healthy eating and sharing this information with advertisers who can provide users with coupons.	27 (60%)	*I like that idea of if somebody wanted to do it so they don't get diabetes. Upload any pictures of the food you eat and get tips about what to do.* *People have to actively want to use it and those individuals may want to receive grocery store coupons that are in line with the goals of losing weight, right? So I don't have a negative reaction to that.*	*I would worry that in the future or – I think it's already possible to – for them to sell the data to another company that could reverse engineer the data and result in higher insurance premiums for individuals and that sort of thing.* *…Nothing would prevent me from taking a picture of the salad that my wife is eating and post that as my dinner, while I'm – in reality, I'm eating a big, thick juicy steak and baked potato…It seems like a lot of effort to go through, and I'm not sure what the benefit's going to be.*	*They have to disclose exactly who they're sharing that information with outside of their company and the advertisers and what specific information they're sharing with those advertisers and the other third parties that may be involved.* *I would think there would have to be a government mandate for a lot of transparency, including specific fines for misusing or not protecting the data.*

[a]Responses by participants to each scenario were coded as favorable, unfavorable, or neutral.

Scenario 1

Scenario 1 was a follows: *A health insurance company accesses data on consumers' grocery store purchases to identify individuals at high risk for developing diabetes and sends them healthy eating tips.*

Overall, 20 out of 45 participants (44%) rated this program favorably. Participants liked the idea they could receive recommendations to improve their diet or management of their health conditions. Others thought the program could address population-level health problems in the United States or reduce insurance costs for those making healthy food choices. Participants were concerned by the motive of health insurers running such a program. They worried about losing insurance coverage or paying higher premiums based on their grocery purchases. Participants described protections they desired, including limits on other uses of data collected for this program, destruction of data after the program was complete, and opt-in and consent procedures for interested participants.

Scenario 2

Scenario 2 was as follows: *A doctor's office monitors internet search data of their patients and contacts individuals who search for symptoms of illness in order to provide an early diagnosis.*

Overall, 12 out of 45 participants (27%) rated this program favorably. Participants identified several beneficial elements of this program, including the prevention of illness, possible savings on health care costs, convenience of care that could be provided remotely particularly for those with barriers to face-to-face care, and additional support from clinical providers for overall health. Additionally, participants liked that this scenario could provide patients with better quality information, preventing inaccurate self-diagnoses using web searches. Some participants felt that having a doctor access their searches would be too invasive, and they would rather conduct personal research and contact their provider if they had concerns. Some felt uncomfortable with the idea that their data could potentially be used for targeted drug or treatment advertising.

Participants emphasized the negative aspects of this scenario. Many were concerned about inadequate protections for their internet search data. They asked for full transparency and limits around the uses of their data, as well as personal data access. Participants also generally preferred that this program be opt-in and wanted assurances that there would be adequate data security safeguards in place.

Scenario 3

Scenario 3 was as follows: *University researchers access the smartphone location and medical records of patients to study risk factors for cancer.*

Overall, 35 out of 45 participants (78%) rated this program favorably. Participants saw benefit in participating in any program that could lead to cancer prevention, often citing personal or family history as motivators for participation. Some viewed this scenario more favorably because consent is required for research participation. Some participants who did not approve felt that the data sharing required was an invasion of their privacy. In particular, some participants indicated that they were not comfortable with researchers having access to their medical records.

Participants suggested that researchers obtain additional consent to have their location tracked and that they have the ability to turn off their location sharing at will. Participants also indicated the importance of transparency, specifically that the location information was secure, was not sold to other companies, and was used solely for research purposes. Finally, participants desired data to be deidentified.

Scenario 4

Scenario 4 was as follows: *A technology company creates an app that tracks users' locations and asks them to upload photos of their meals with the intention of encouraging healthy eating and sharing this information with advertisers who can provide users with coupons.*

Overall, 27 out of 45 participants (60%) rated this program favorably. Those who responded favorably to this program thought it would help them to eat healthier and prevent diabetes, and many wanted to receive coupons for using the app. A few appreciated the voluntary or opt-in nature of the app. Others were mistrustful of the motives and felt that only the company or insurers would benefit.

Many participants did not want their information to be sold to advertising companies. Some would not participate because of the inconvenience of having to use the app consistently or the nuisance of receiving advertisements. Others had privacy concerns and worried about downstream use of their information for other purposes. For many, the only desired protection was the ability to either opt-in or opt-out of the app. Finally, participants discussed the importance of anonymity and the ability to use the app without advertisers or third parties identifying them.

Discussion

Principal Findings

In a broadly representative sample of US consumers, we examined attitudes and knowledge about digital health privacy. We identified several key findings. First, participants generally did not draw connections between the data they leave behind when they use digital technology and what it may reveal about their health. Second, participants struggled to weigh harms and benefits of data collection and use. Third, while preferences varied, participants generally wanted some protections in place for health-relevant consumer digital data.

Prior work has demonstrated that consumers value privacy but nonetheless continue to use digital technologies that routinely compromise that privacy [19]. Turow et al argued that this behavior should not be interpreted as consumers being unconcerned with privacy, but rather a sense of resignation that threats to privacy are unavoidable [20]. Our findings suggest that consumers often recognize that information is shared when they use digital technology but are often unaware of the extent of use or what can be learned about them from their data. Obar et al demonstrated that simply presenting consumers with more information will not necessarily lead to more informed consumers or yield higher quality decision making. In an experiment simulating registration with a social networking program, they found that consumers viewed lengthy privacy policies as a nuisance, felt overburdened by privacy statements, and missed "gotcha" provisions that were extreme (eg, many consented to give up their first-born child) [21].

Given that many consumers rank health privacy above other areas [22], it is concerning that most participants in our study could not draw connections between consumer digital data and inferences about their health. Broadly available analytic tools allow for the health status to be inferred from a wide range of data sources including smart devices in the home, language analysis from social media, financial spending habits, and internet search patterns [7,8,23]. Even when health applications were made explicit in the study scenarios, the tradeoffs between threats to privacy and benefits of use remained difficult for our study participants to reconcile. The difficulty assessing tradeoffs is likely compounded by the absence of limits on data use and the uncertain risks posed. Conger et al have argued that the relationships between consumers and businesses have become too complex for the individual consumer to assess given the involvement of third-party data brokers and open-ended uses of data [24].

Facing vignettes about commercial users, participants often raised concerns about economic harms (eg, higher insurance premiums and threats to employment) as well as unbridled secondary uses of their data. When considering vignettes with researchers and doctors as the data users, participants were often concerned with reputational harms (eg, sharing of embarrassing or unflattering information) [25]. Overall, participants responded much more favorably to some scenarios over others suggesting that consumers apply a nuanced evaluation of the use, including whether it is beneficial to themselves, the type of data that is being used, who is using their data, and the overall benefit of the use.

Participants sought privacy safeguards to reconcile the tradeoffs between health benefits and privacy. Many of the safeguards raised mirror those that have been part of the EU General Data Protection Regulation, including consent for data collection and use (opt-in), explicit statements of what data would be used for and by whom, limits on data transfer and selling, and data security requirements [26]. Participants were overwhelmingly in favor of protections for their digital health data. However, prior work points to practical limitations on the effectiveness of some of these approaches. For example, consumers often click through overly complex privacy agreements, suggesting a vastly simplified approach would be needed [20]. Policymakers might consider varying privacy protections based on the associated risks and benefits of different uses or curtail downstream transfers and aggregation of data that are far removed from the original point of collection.

Limitations

This study has several limitations. Digital privacy is a topic with growing relevance and has been increasingly highlighted in popular media, which could shape consumers' attitudes. Events just prior to interviews may have heavily influenced participants' knowledge and perceptions. Therefore, the interviews reflect the time period (2018-2019) in which they were conducted. Despite efforts to recruit consumers with a wide range of characteristics, consumers who were willing to respond to and complete a request for an interview may have characteristics or attitudes not shared by those who chose not to participate. Social desirability bias may limit the credibility of participants' expressed views. Despite these limitations, the interviews show a diverse array of attitudes regarding digital privacy.

Conclusions

This study reveals how individual consumers wish to protect their own health information privacy and how little they are aware of the threats to that privacy from their conventional behaviors. Digital personal health information is now derived from consumer engagements as well as medical engagements. To the extent we need protections for personal health privacy, those protections need to extend beyond their current reach. Our study reinforces the finding that a purely market-based approach to privacy that depends on rational consumers making decisions with full information relies on faulty assumptions [27]. Future debates in the United States may need to focus on whether more robust privacy standards (similar to the European Union's General Data Protection Regulation) are needed to protect consumer digital data from entities beyond health care systems [28,29].

Acknowledgments

This research was supported by the National Human Genome Research Institute/National Institutes of Health (NIH) (5R01HG009655-04). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. We thank Ashley Lewson for her role in conducting and coding interviews.

Conflicts of Interest

DAA is a partner at VAL Health. The other authors have no conflicts to declare.

Multimedia Appendix 1
Interview guide.
[DOCX File , 27 KB - jmir_v23i6e29395_app1.docx]

References

1. Digital 2020: 3.8 Billion People Use Social Media. We Are Social. URL: https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media [accessed 2021-02-12]
2. Kaplan B. Seeing Through Health Information Technology: The Need for Transparency in Software, Algorithms, Data Privacy, and Regulation. SSRN Journal 2020 Oct 09;7(1):1-18. [doi: 10.2139/ssrn.3672395]

3. McIver DJ, Hawkins JB, Chunara R, Chatterjee AK, Bhandari A, Fitzgerald TP, et al. Characterizing Sleep Issues Using Twitter. J Med Internet Res 2015 Jun 08;17(6):e140-e140 [FREE Full text] [doi: 10.2196/jmir.4476] [Medline: 26054530]
4. Card C. How Facebook AI Helps Suicide Prevention. Facebook. 2018. URL: https://about.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/ [accessed 2021-01-13]
5. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: WebSci '13: Proceedings of the 5th Annual ACM Web Science Conference. 2013 May Presented at: 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris, France p. 47-56. [doi: 10.1145/2464464.2464480]
6. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018 Oct 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]
7. Grande D, Luna Marti X, Feuerstein-Simon R, Merchant RM, Asch DA, Lewson A, et al. Health Policy and Privacy Challenges Associated With Digital Technology. JAMA Netw Open 2020 Jul 01;3(7):e208285-e208285 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8285] [Medline: 32644138]
8. All Your Data Is Health Data. The New York Times. 2019. URL: https://www.nytimes.com/2019/08/13/opinion/health-data.html [accessed 2020-02-11]
9. Sonn JW. Coronavirus: South Korea's success in controlling disease is due to its acceptance of surveillance. The Conversation. 2020. URL: http://theconversation.com/coronavirus-south-koreas-success-in-controlling-disease-is-due-to-its-acceptance-of-surveillance-134068 [accessed 2021-02-20]
10. Stokes DC, Andy A, Guntuku SC, Ungar LH, Merchant RM. Public Priorities and Concerns Regarding COVID-19 in an Online Discussion Forum: Longitudinal Topic Modeling. J Gen Intern Med 2020 Jul 12;35(7):2244-2247 [FREE Full text] [doi: 10.1007/s11606-020-05889-w] [Medline: 32399912]
11. Guntuku SC, Sherman G, Stokes DC, Agarwal AK, Seltzer E, Merchant RM, et al. Tracking Mental Health and Symptom Mentions on Twitter During COVID-19. J Gen Intern Med 2020 Sep 07;35(9):2798-2800 [FREE Full text] [doi: 10.1007/s11606-020-05988-8] [Medline: 32638321]
12. Schechner S, Winkler R. Here's How Apple and Google Plan to Track the Coronavirus Through Your Phone. The Wall Street Journal. 2020. URL: https://www.wsj.com/articles/heres-how-apple-and-google-plan-to-track-the-coronavirus-through-your-phone-11586618075 [accessed 2020-11-20]
13. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug 07;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
14. Auxier B, Rainie L, Anderson M, Perrin A, Kumar M, Turner E. Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/ [accessed 2020-09-20]
15. KnowledgePanel: A Methodological Overview. IPSOS. URL: https://www.ipsos.com/sites/default/files/ipsosknowledgepanelmethodology.pdf [accessed 2020-01-18]
16. Institute of Medicine, Committee on Regional Health Data Networks, Donaldson MS, Lohr KN. Health Data in the Information Age: Use, Disclosure, and Privacy. Washington, DC: The National Academies Press; 1994.
17. NVivo Qualitative Data Analysis Software. QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2020-03-20]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan 1;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Madden M, Rainie L. Americans' Views About Data Collection and Security. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/05/20/americans-views-about-data-collection-and-security/ [accessed 2021-02-11]
20. Turow J, Hennessy M. The Tradeoff Fallacy: How Marketers are Misrepresenting American Consumers and Opening Them Up to Exploitation. SSRN Journal 2015 Jun 1:1-24. [doi: 10.2139/ssrn.2820060]
21. Obar JA, Oeldorf-Hirsch A. The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services. Information, Communication & Society 2018 Jul 03;23(1):128-147. [doi: 10.1080/1369118x.2018.1486870]
22. Madden M. Americans Consider Certain Kinds of Data to be More Sensitive than Others. Pew Research Center. 2014. URL: https://www.pewresearch.org/internet/2014/11/12/americans-consider-certain-kinds-of-data-to-be-more-sensitive-than-others/ [accessed 2021-02-11]
23. Fussell S. Google's Totally Creepy, Totally Legal Health-Data Harvesting. The Atlantic. 2019. URL: https://www.theatlantic.com/technology/archive/2019/11/google-project-nightingale-all-your-health-data/601999/ [accessed 2021-11-02]
24. Conger S, Pratt J, Loch K. Personal information privacyemerging technologies. Information Systems Journal 2013;23(5):401-417. [doi: 10.1111/j.1365-2575.2012.00402.x]
25. Prasad A, Sorber J, Stablein T, Anthony D, Kotz D. Understanding sharing preferences and behavior for mHealth devices. In: WPES '12: Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society. 2012 Oct Presented at: 2012

ACM Workshop on Privacy in the Electronic Society; October 15, 2012; Raleigh, NC p. 117-128. [doi: 10.1145/2381966.2381983]

26. Voigt P, von dem Bussche A. The EU General Data Protection Regulation (GDPR). New York, NY: Springer Publishing; 2017.
27. Park YJ. Does a market-approach to online privacy protection result in better protection for users? The Policy and Internet Blog. 2015. URL: https://blogs.oii.ox.ac.uk/policy/does-a-market-approach-to-online-privacy-protection-result-in-better-protection-for-users/ [accessed 2021-03-20]
28. Orlando AW, Rosoff AJ. The New Privacy Crisis: What's Health Got to Do with It? Am J Med 2019 Feb;132(2):127-128. [doi: 10.1016/j.amjmed.2018.09.033] [Medline: 30367855]
29. Gostin LO, Halabi SF, Wilson K. Health Data and Privacy in the Digital Era. JAMA 2018 Jul 17;320(3):233-234. [doi: 10.1001/jama.2018.8374] [Medline: 29926092]

Edited by G Eysenbach; submitted 05.04.21; peer-reviewed by E Oghogho, P Swire; comments to author 26.04.21; revised version received 10.05.21; accepted 16.05.21; published 09.06.21.

Please cite as:
Grande D, Luna Marti X, Merchant RM, Asch DA, Dolan A, Sharma M, Cannuscio CC
Consumer Views on Health Applications of Consumer Digital Data and Health Privacy Among US Adults: Qualitative Interview Study
J Med Internet Res 2021;23(6):e29395
URL: https://www.jmir.org/2021/6/e29395
doi:10.2196/29395
PMID:34106074

Original Paper

Public Interest in Immunity and the Justification for Intervention in the Early Stages of the COVID-19 Pandemic: Analysis of Google Trends Data

Jinhee Lee[1*], MD; Yunna Kwan[1,2*], MA; Jun Young Lee[3], MD; Jae Il Shin[4], MD; Keum Hwa Lee[4], MD; Sung Hwi Hong[4], MD; Young Joo Han[5], MD; Andreas Kronbichler[6], MD; Lee Smith[7], PhD; Ai Koyanagi[8,9], PhD; Louis Jacob[8,10], MD; SungWon Choi[2], PhD; Ramy Abou Ghayda[11,12], PhD; Myung-Bae Park[13], PhD

[1]Department of Psychiatry, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea
[2]Department of Psychology, Duksung Women's University, Seoul, Republic of Korea
[3]Department of Nephrology, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea
[4]Department of Pediatrics, Yonsei University College of Medicine, Seoul, Republic of Korea
[5]Department of Pediatrics, Samsung Changwon Hospital, Sungkyunkwan University School of Medicine, Changwon, Republic of Korea
[6]Department of Internal Medicine IV, Nephrology and Hypertension, Medical University Innsbruck, Innsbruck, Austria
[7]The Cambridge Centre for Sport and Exercise Sciences, Anglia Ruskin University, Cambridge, United Kingdom
[8]Parc Sanitari Sant Joan de Déu/Centro de Investigación Biomédica en Red de Salud Mental, Universitat de Barcelona, Barcelona, Spain
[9]Catalan Institution for Research and Advanced Studies, Barcelona, Spain
[10]Faculty of Medicine, University of Versailles Saint-Quentin-en-Yvelines, Paris, France
[11]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States
[12]Urology Institute, University Hospitals, Case Western Reserve University, Cleveland, OH, United States
[13]Department of Gerontology Health and Welfare, Pai Chai University, Daejeon, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Myung-Bae Park, PhD
Department of Gerontology Health and Welfare
Pai Chai University
155-40, Baejae-ro, Seo-gu
Daejeon, 35345
Republic of Korea
Phone: 82 425205037
Email: parkmb@pcu.ac.kr

Abstract

Background: The use of social big data is an important emerging concern in public health. Internet search volumes are useful data that can sensitively detect trends of the public's attention during a pandemic outbreak situation.

Objective: Our study aimed to analyze the public's interest in COVID-19 proliferation, identify the correlation between the proliferation of COVID-19 and interest in immunity and products that have been reported to confer an enhancement of immunity, and suggest measures for interventions that should be implemented from a health and medical point of view.

Methods: To assess the level of public interest in infectious diseases during the initial days of the COVID-19 outbreak, we extracted Google search data from January 20, 2020, onward and compared them to data from March 15, 2020, which was approximately 2 months after the COVID-19 outbreak began. In order to determine whether the public became interested in the immune system, we selected *coronavirus*, *immune*, and *vitamin* as our final search terms.

Results: The increase in the cumulative number of confirmed COVID-19 cases that occurred after January 20, 2020, had a strong positive correlation with the search volumes for the terms *coronavirus* (R=0.786; P<.001), *immune* (R=0.745; P<.001), and *vitamin* (R=0.778; P<.001), and the correlations between variables were all mutually statistically significant. Moreover, these correlations were confirmed on a country basis when we restricted our analyses to the United States, the United Kingdom, Italy, and Korea. Our findings revealed that increases in search volumes for the terms *coronavirus* and *immune* preceded the actual occurrences of confirmed cases.

Conclusions: Our study shows that during the initial phase of the COVID-19 crisis, the public's desire and actions of strengthening their own immune systems were enhanced. Further, in the early stage of a pandemic, social media platforms have a high potential for informing the public about potentially helpful measures to prevent the spread of an infectious disease and provide relevant information about immunity, thereby increasing the public's knowledge.

(J Med Internet Res 2021;23(6):e26368) doi:10.2196/26368

KEYWORDS

COVID-19; social big data; infodemiology; infoveillance; social listening; immune; vitamin; big data; public interest; intervention; immune system; immunity; trends; Google Trends; internet; digital health; web-based health information; correlation; social media; infectious disease

Introduction

Following the onset of an infectious pneumonia that could be traced to Wuhan, China, in 2019, the World Health Organization (WHO) announced that the cause of the pneumonia was a new type of coronavirus. The International Committee on Taxonomy of Viruses named it SARS-CoV-2 on on January 11, 2020, and the infectious pneumonia was referred to as COVID-19. As of December 2020, the virus has spread globally; the cumulative number of people infected with SARS-CoV-2 has reached approximately 67 million, and approximately 1,530,000 have died from the virus [1].

Since the year 2000, there have been attempts to use big data to better understand and solve challenges related to public health. Indeed, in the context of health care, the use of big data is an important emerging trend that is only likely to increase in importance over time. The Pillbox project of the National Laboratory of Medicine is one of the most representative examples of health care that uses big data to predict real-world situations. The Pillbox project is a service that provides information such as how to take various pills that are not well known by the public and their potential side effects. The use of such information not only improves consumer convenience but also saves health care expenditures with regard to economics through the gathering of information to estimate statistics, such as the causes of disease outbreaks, rates of spread, and distribution and regional growth [2,3]. Social big data refers to large volumes of data that relate to people or describe people's behaviors and technology-mediated social interactions in the digital realm [4]. Internet search volume is a valid and effective tool for detecting public attention [5]. One of the most representative examples of using social big data for disease prediction is Google's flu forecast. In 2009, Google predicted the spread of the flu 7 to 10 days earlier than the Centers for Disease Control and Prevention (CDC) based on users' search records for the flu [6]. This prediction was confirmed in countries around the world, including South Korea, India, and China [7-9]. Predictive analyses that use social big data, which are not easily accessible in reality, help with making relatively accurate predictions of real-world events. In recent times, attempts have been made to further increase the precision of disease predictions [10,11]. Furthermore, predictive analyses were also used for the real-time monitoring of SARS (severe acute respiratory syndrome) and its transmissibility and for the identification of the natural history of emerging pathogens, such as the Ebola virus [12,13].

The spread of infectious diseases has led to increased interest in the human immune system among the public. During a pandemic, the public often becomes interested in how to strengthen their immune systems. Furthermore, the act of purchasing products that may strengthen one's immune system is popular, even without clear evidence that shows whether these supplements are effective. However, there is almost no scientific proof for confirming this trend, and there is a need to confirm that such interest results in people actually searching more areas of the internet for products related to strengthening their immune system. Eventually, people are likely to respond to this lack of evidence by relying on existing common sense when it comes to maintaining health and hygiene. In this regard, a health literacy approach is needed during a crisis such as the COVID-19 pandemic. At the beginning of a pandemic, the public's response to an infectious disease is the most sensitive, and people have a burgeoning interest in infectious diseases during this period. After the passage of a certain period of time, exhaustion and the adjustment to the social situation results in a relative decline in this interest [14].

Our study aims to analyze the public's interest in the global pandemic and correlate internet search volumes for COVID-19–related terms with the public's interest in the immune system and vitamins.

Methods

Data Extraction

To assess the public's interest in infectious diseases during the initial days of the COVID-19 outbreak, we extracted Google Trends search data from January 20, 2020, onward and compared them to data from March 15, 2020, which was approximately 2 months after the COVID-19 outbreak began. Apart from in a few countries such as China and Japan, the Google search engine has an overwhelming dominance of market shares [15]. Google provides data on search word volumes through Google Trends, which provides usable data for specific countries and time periods. In this study, our unit of data was extracted from data on total worldwide numbers. In addition, to account for trends by country, we selected the United States and the United Kingdom for analysis, as these countries are the most representative countries of the English-speaking world that use the Google search engine. Moreover, we also studied Italy, which was the first European country to experience an explosion in the number of COVID-19 cases. The market share of Google in Italy in January 2020 was 98.9% [16]. Finally, we studied South Korea, which had the second highest COVID-19 spread rate in Asia and the fourth highest spread rate in the world at the time the Google Trends search was performed. On the date that the data were gathered, South Korea was ranked fourth in the world for COVID-19 case numbers, with 8000 cases. In South Korea, the domestic portal site Naver (Naver Corporation) has a much higher market share than Google. Therefore, additional data supplied by Naver Data Lab were used for cross-verification.

Primary Keywords

We used the terms *covid*, *corona*, and *coronavirus* as search terms. In order to determine which of these words were the most commonly used, the search volume of each word was ascertained. According to our preliminary analysis of the data provided by Google Trends, the term *coronavirus* was searched twice more often than *corona* and 5 times more often than *covid*. Therefore, we selected *coronavirus* as our final search term.

Searches Regarding Immunity and Products That Might Strengthen Immunity

In order to determine whether the public became interested in the immune system, we selected *immune* as a keyword. The term *immune* was strongly related to the term *immunity* and had an almost 2-fold higher search volume. Therefore, for the United States, United Kingdom, and Italy, the search term *immune* was studied, and in the case of South Korea, the term *myeonyeok*, which means “immunity” in Korean, was studied.

Following the outbreak of COVID-19, the public’s interest in products that can enhance one’s immune system increased dramatically. The Guardian in the United Kingdom and the Washington Post in the United States have both reported on the rising demand for vitamins and other products [17]. Vitamins are the most familiar to the public, are usually available over the counter, and are the most representative product that enhances the immune system. Therefore, in order to analyze the public’s actual interest in immune system strength, we selected the term *vitamin* as a keyword. For the United States and the United Kingdom, the term *vitamin* was used. For Italy, the term *vitamina* was used. Finally, for South Korea, the Korean-language spelling of the term *vitamin* was used.

Data and Statistical Analysis

Search volumes were calculated for each period (1 week), and the search volume of the period with the highest search volume was 100, which normalized every other search volume value from 0 to 100. The unit of extraction was 1 week. In order to account for the spread of COVID-19, we used the cumulative number of confirmed cases. The official announcements made by the WHO and announcements from each country’s respective government were the most accurate indices for tracking the spread of SARS-CoV-2. The number of cumulative, confirmed cases was verified with data downloaded from Our World in Data [18] and additional data from COVID-19 situation reports issued by the WHO [1]. We conducted an analysis of the correlation between the number cumulative, confirmed COVID-19 cases and search terms. Up until the middle of January 2020, before COVID-19 had begun to spread on a large scale, the figure for confirmed cases and search volume was close to 0. In other words, the actual figure was 0 or 1; therefore, a correlation analysis of the entire year could lead to errors. As such, we selected countries that experienced some of the earliest COVID-19 outbreaks as our target countries. Moreover, we analyzed approximately 2 months’ worth of data, beginning from when COVID-19 cases first began to emerge globally during the week of January 20, 2020, up until March 15, 2020.

Data Availability

Data can be downloaded from the Google Trends website. If the processed data are needed, the authors can be contacted to request the data.

Results

Up until the middle of January 2020, the number of confirmed COVID-19 cases remained extremely low worldwide. Afterward, beginning in the week of January 26, the number of confirmed cases began to increase rapidly, especially in China. The search volume for the term *coronavirus* began to increase earlier—beginning in the week of January 12—and the search volume increased rapidly during the week of February 16. For the past year, the search volume for the term *immune* has consistently been between 10 and 20. After January 12, the search volume increased steadily, and from February 16 onward, the search volume increased rapidly. Throughout 2020, the search volume for the term *vitamin* was at a level that slightly exceeded 50. However, this began to increase after February. The increase in the number of cumulative, confirmed cases of COVID-19 after January 20 had a strong positive correlation with the search volumes for the terms *coronavirus* (R=0.786; P<.001), *immune* (R=0.745; P<.001), and *vitamin* (R=0.778; P<.001), and the correlations between variables were all mutually statistically significant (Figure 1, Table 1).

Figure 1. Worldwide trends in search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the number of cumulative, confirmed COVID-19 cases over the past year (from March 18, 2019, to March 15, 2020).

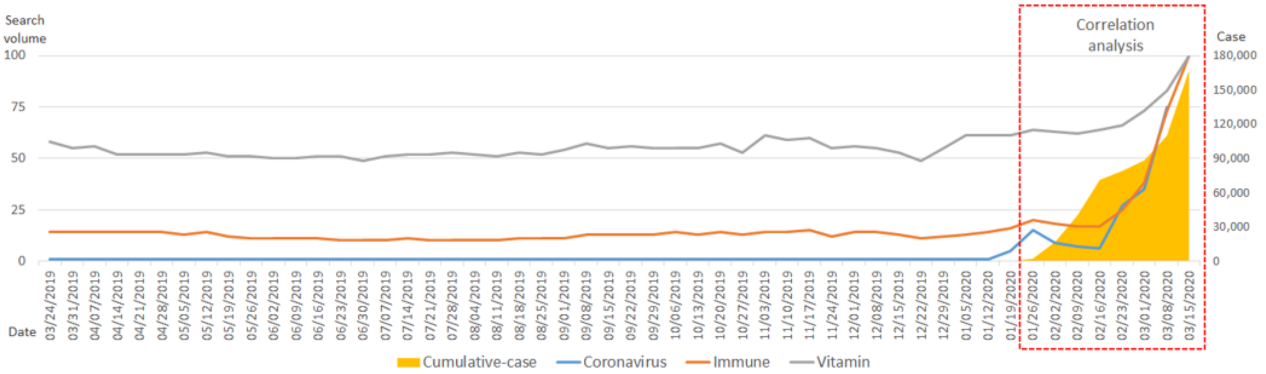


Table 1. Correlations among search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the actual number of cumulative, confirmed COVID-19 cases worldwide (over 56 days).

Variable	Cumulative case numbers	Coronavirus	Immune	Vitamin
Cumulative case numbers				
R	1	0.786	0.745	0.778
P value	—[a]	<.001	<.001	<.001
Coronavirus				
R	0.786	1	0.979	0.932
P value	<.001	—	<.001	<.001
Immune				
R	0.745	0.979	1	0.943
P value	<.001	<.001	—	<.001
Vitamin				
R	0.778	0.932	0.943	1
P value	<.001	<.001	<.001	—

[a] Not applicable.

With respect to the United States, up until February 2020, the number of cumulative, confirmed cases remained low (10 people). This number began increasing in the latter half of February and began increasing rapidly in March. The search volume for the term *coronavirus* began increasing earlier—beginning in the week of January 12—and it began to increase rapidly in the week of February 16. The search volume for the term *immune* began to rapidly increase on February 16. Over the past year, the search volume for the term *vitamin* was at a level that slightly exceeded 50. This number increased slightly after December 2019 and increased more rapidly in March 2020. The increase in the number of cumulative, confirmed cases after January 20 had a strong positive correlation with the search volumes for the terms *coronavirus* (R=0.921; P<.001), *immune* (R=0.890; P<.001), and *vitamin* (R=0.913; P<.001), and the correlations between variables were all mutually statistically significant (Figure 2, Table 2).

Figure 2. Trends in search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the number of cumulative, confirmed COVID-19 cases in the United States over the past year (March 18, 2019, to March 15, 2020).

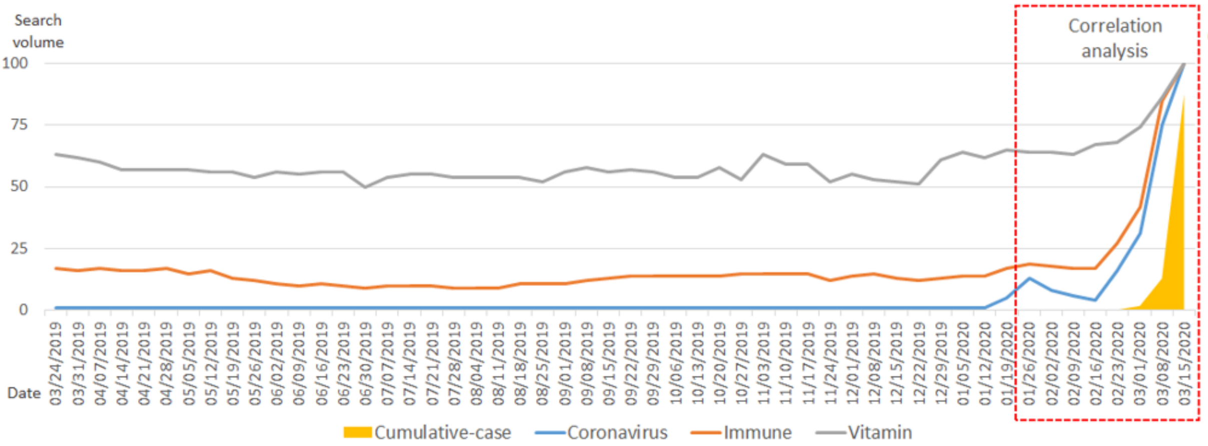


Table 2. Correlations among search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the actual number of cumulative, confirmed COVID-19 cases in the United States (over 56 days).

Variable	Cumulative case numbers	Coronavirus	Immune	Vitamin
Cumulative case numbers				
R	1	0.921	0.890	0.913
P value	—[a]	<.001	<.001	<.001
Coronavirus				
R	0.921	1	0.983	0.952
P value	<.001	—	<.001	<.001
Immune				
R	0.890	0.983	1	0.946
P value	<.001	<.001	—	<.001
Vitamin				
R	0.913	0.952	0.946	1
P value	<.001	<.001	<.001	—

[a] Not applicable.

In the United Kingdom, the number of confirmed cases increased rapidly after March 2020. Increases in search volumes for the terms *coronavirus*, *immune*, and *vitamin* followed a trend that was similar to those observed for the United States and the rest of the world. The number of cumulative, confirmed cases had a strong positive correlation with the search volumes for the terms *coronavirus* (R=0.931; P<.001), *immune* (R=0.962; P<.001), and *vitamin* (R=0.801; P<.001), and the correlations between variables were mutually statistically significant (Figure 3, Table 3).

Figure 3. Trends in search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the number of cumulative, confirmed COVID-19 cases in the United Kingdom over the past year (March 18, 2019, to March 15, 2020).

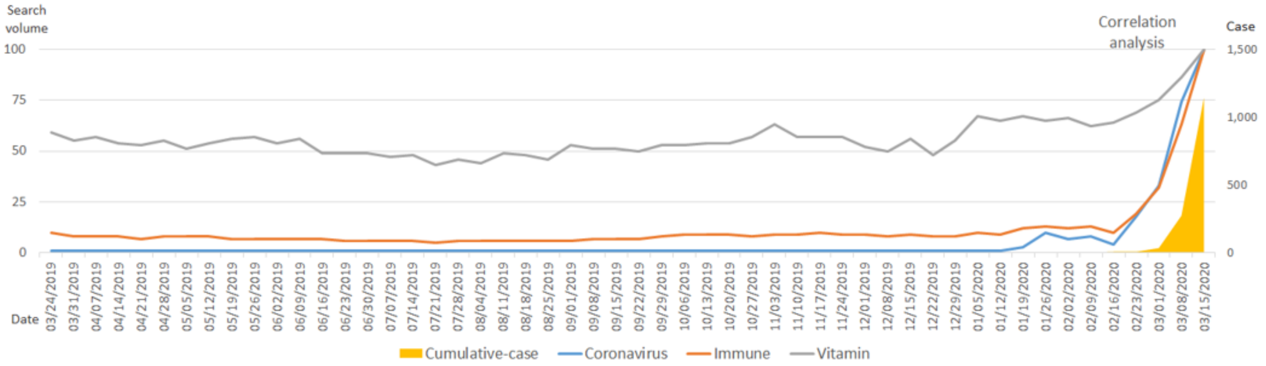


Table 3. Correlations among search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the actual number of cumulative, confirmed COVID-19 cases in the United Kingdom (over 56 days).

Variable	Cumulative case numbers	Coronavirus	Immune	Vitamin
Cumulative case numbers				
R	1	0.931	0.962	0.801
P value	—[a]	<.001	<.001	<.001
Coronavirus				
R	0.931	1	0.984	0.842
P value	<.001	—	<.001	<.001
Immune				
R	0.962	0.984	1	0.847
P value	<.001	<.001	—	<.001
Vitamin				
R	0.801	0.842	0.847	1
P value	<.001	<.001	<.001	—

[a] Not applicable.

In Italy, the number of confirmed cases started to rise in the second half of February 2020, with case numbers increasing dramatically in March. The search volumes for the terms *coronavirus* and *immune* began to increase slightly in the week of January 26, and they began to increase dramatically in the middle of February. The search volume for the term *vitamin* first began to increase rapidly in the middle of February, but afterward, it repeatedly decreased and increased. The number of cumulative, confirmed cases had a positive correlation with the search terms *coronavirus* (*R*=0.600; *P*<.001), *immune* (*R*=0.763; *P*<.001), and *vitamin* (*R*=0.474; *P*<.001), and the correlations between variables were all mutually statistically significant (Figure 4, Table 4).

Figure 4. Trends in search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the number of cumulative, confirmed COVID-19 cases in Italy over the past year (March 18, 2019, to March 15, 2020).

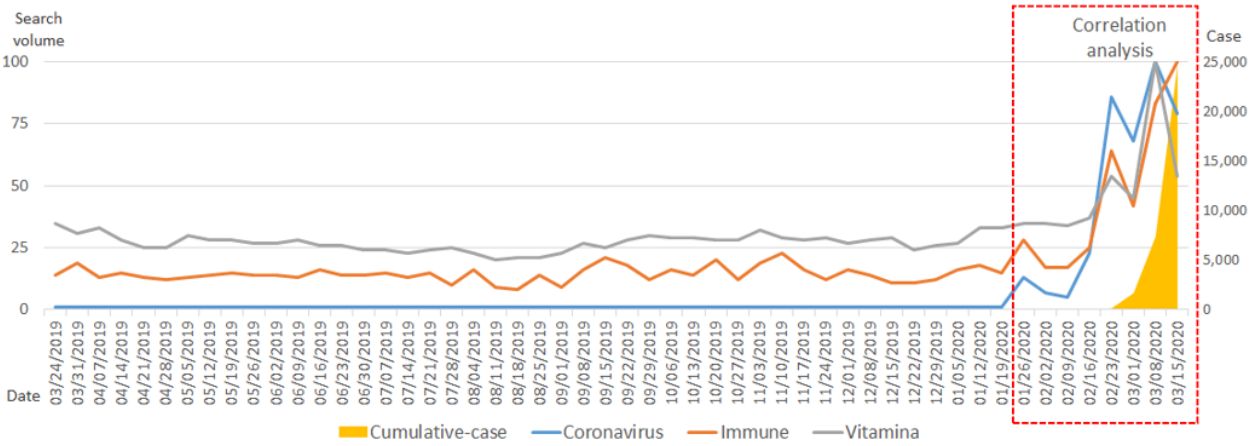


Table 4. Correlations among search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the actual number of cumulative, confirmed COVID-19 cases in Italy (over 56 days).

Variable	Cumulative case numbers	Coronavirus	Immune	Vitamin
Cumulative case numbers				
R	1	0.931	0.962	0.801
P value	—[a]	<.001	<.001	<.001
Coronavirus				
R	0.931	1	0.984	0.842
P value	<.001	—	<.001	<.001
Immune				
R	0.962	0.984	1	0.847
P value	<.001	<.001	—	<.001
Vitamin				
R	0.801	0.842	0.847	1
P value	<.001	<.001	<.001	—

[a] Not applicable.

In South Korea, the number of cumulative, confirmed cases increased rapidly in the middle of February. The search volume for the term *coronavirus* increased earlier—beginning in the week of January 19. After decreasing for a brief period, this search volume again increased rapidly in the week of February 9 and again decreased in March. The search volume for the term *immune* was highest during the first week of March, whereas the search volume for the term *vitamin* was highest during the week of March 15. However, unlike in other countries, the search volumes for these terms did not rapidly increase. The increase in the number of cumulative, confirmed cases had a statistically significant positive correlation with the search volumes for the terms *coronavirus* (*R*=0.359; *P*=.007) and *vitamin* (*R*=0.637; *P*<.001); however, its correlation with the search volume for the term *immune* (*R*=0.254; *P*=.06) was not statistically significant. When analyzing data from Naver, the search volume for the term *coronavirus* increased from the middle of February onward, and the search volumes for the terms *immune* and *vitamin* increased slightly in the beginning of January and increased again in the second half of January and from February 16 onward. All three search terms reached their peak search volumes in the beginning of March, and thereafter, these search volumes began trending downward. When compared with data from Google Trends, the results were similar, and the increasing trend in search volumes for the terms *immune* and *vitamin* that appeared after the occurrence of confirmed cases was slightly clearer (Figure 5, Table 5).

Figure 5. Trends in search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the number of cumulative, confirmed COVID-19 cases in South Korea over the past year (March 18, 2019, to March 15, 2020).

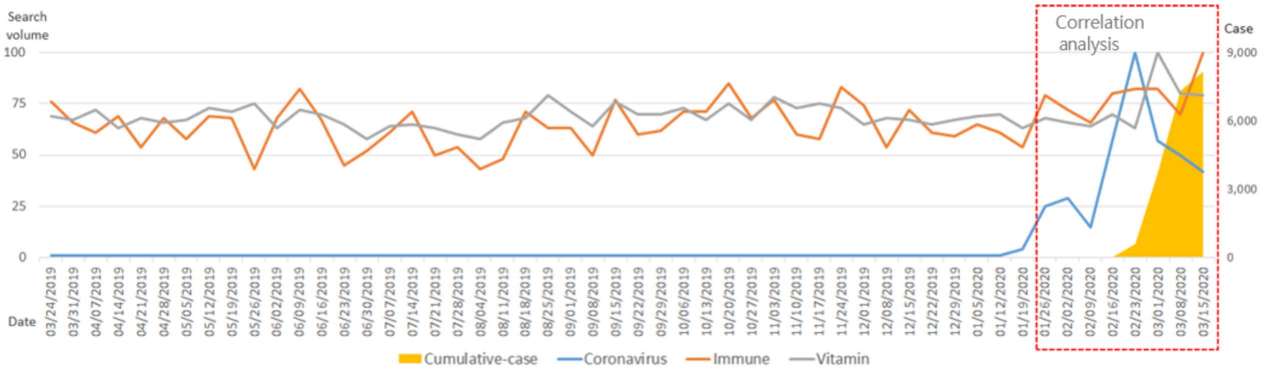


Table 5. Correlations among search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and and the actual number of cumulative, confirmed COVID-19 cases in South Korea (over 56 days).

Variable	Cumulative case numbers	Coronavirus	Immune	Vitamin
Cumulative case numbers				
R	1	0.359	0.254	0.637
P value	—[a]	.007	.06	<.001
Coronavirus				
R	0.359	1	0.157	0.242
P value	.007	—	.25	.07
Immune				
R	0.254	0.157	1	0.368
P value	.06	.25	—	.005
Vitamin				
R	0.637	0.242	0.368	1
P value	<.001	.07	.005	—

[a] Not applicable.

Discussion

During the COVID-19 pandemic and other pandemics, educational approaches to health and health care are an important problem from the perspective of public health. This study analyzed the increase in the public's interest in COVID-19 by using social media big data and attempted to gain knowledge concerning what interventions must be implemented from a health and health care perspective. Our findings revealed that increases in search volumes for the terms *coronavirus* and *immune* preceded the actual occurrences of confirmed COVID-19 cases. Moreover, search volumes increased gradually before increasing rapidly at the same time as when a rapid increase in the number of confirmed cases was reported. Search volumes for the term *vitamin* also increased concomitantly. This occurred at the same time as the increase in the number of confirmed COVID-19 cases. Although there were slight differences in the increasing trends among countries, worldwide trends coincided with each other in general. With respect to South Korea, although Google search results did correspond with worldwide trends in a characteristic manner, the data from the domestic portal Naver corresponded more closely with worldwide trends (Figure 6, Table 6).

Figure 6. Trends in search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the number of cumulative, confirmed COVID-19 cases in South Korea over the past year (March 18, 2019, to March 15, 2020) based on data from the Naver portal.

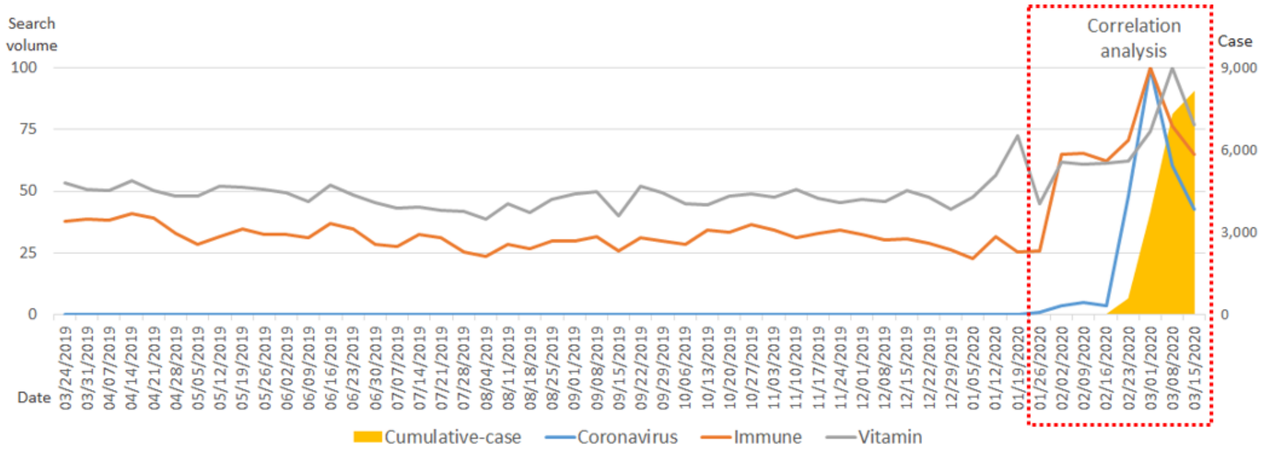


Table 6. Correlations among search volumes for the terms *coronavirus*, *immune*, and *vitamin*, and the actual number of cumulative, confirmed COVID-19 cases in South Korea based on data from the Naver portal (over 56 days).

Variable	Cumulative case numbers	Coronavirus	Immune	Vitamin
Cumulative case numbers				
R	1	0.482	0.223	0.629
P value	—[a]	<.001	.01	<.001
Coronavirus				
R	0.482	1	0.742	0.563
P value	<.001	—	<.001	<.001
Immune				
R	0.223	0.742	1	0.652
P value	.01	<.001	—	<.001
Vitamin				
R	0.629	0.563	0.652	1
P value	<.001	<.001	<.001	—

[a] Not applicable.

By using search records from its search engine, Google was able to predict the 2009 flu pandemic before the CDC. Similarly, this study also found that the increase in the volume of related search terms preceded an increase in the number of confirmed COVID-19 cases, thereby confirming the validity of using social big data for predicting real-world events. However, there are several possible interpretations for the increased volume of certain search terms. First, this may reflect a rising concern about infectious diseases, such as worries about the outbreak of a disease and demands for knowledge. Second, the results of this study confirmed that in addition to the search volume for the term *coronavirus* increasing, the search volume for the terms *immune* and *vitamin* also similarly increased. This can also be interpreted as an act of disease prevention, which reflects a desire to control situations and prevent disease. During the period of the SARS pandemic, as an act of disease prevention, the number of searches for information on the internet and consultations with experts by people in the United States and Canada who were highly concerned about the disease increased [19]. It is known that increased interest and concern about a disease has a very close relationship with disease prevention measures, and it appears that during a pandemic, increased concern about disease is intimately related to handwashing, personal hygiene and related activities, mask wearing, and the avoidance of public places [19].

In particular, the fact that an increase in search volume preceded the spread of COVID-19 implies that before the complete outbreak of a pandemic, the public has a tendency to be concerned about a disease and engage in corresponding prevention activities. COVID-19 symptoms do not present until 5 to 6 days after SARS-CoV-2 infection, and the incubation period can be as long as 14 days [20]. Taking this into account, the results of this study—those showing that an increase in search volumes directly precedes a rapid increase in cases—reflects the coping mechanisms of a public that has already been exposed to COVID-19 and has been experiencing early warning symptoms. Regardless, before the onset of a full-blown pandemic, supplying prior education and information on health care and health should be the central task of public health departments during these situations.

It cannot be ruled out that an increase in search volumes related to COVID-19 reflects an increase in simple interest. Even if this is the case, it is important to place importance on the dissemination of proper knowledge. The control of a disease's outbreak cannot rely solely on the efforts of public health specialists. Coping behaviors and prevention measures that are achieved through an increase in knowledge among the public are critically needed [21,22]. In particular, behavioral measures such as social distancing are an effective strategy for coping with an epidemic. However, it is only possible to achieve this through the cooperation of a well-informed public [22]. It is certain that Wuhan residents who moved to other cities without knowing whether they were infected with SARS-CoV-2 contributed to the early spread of COVID-19 [23]. In addition, homeless people and those of similar social classes that have little access to medical information have been identified as potentially dangerous groups that might be transmitters of the disease [24]. Therefore, the propagation of health care information through mass communication methods is emerging as an issue that is more important than ever before. Therefore, in an epidemic or pandemic situation, a variety of strategies are needed to use social media and the internet to alleviate people's fears and discomfort and increase the public's awareness of proper health measures.

Alongside the COVID-19 pandemic, not only did the search volumes of the terms *immune* and *vitamin* increase, the number of news reports related to COVID-19 that contained these keywords increased as well (Multimedia Appendix 1). Moreover, in addition to an increase in interest in food products such as elderberry, which are known to be beneficial in enhancing the immune system, because these news reports were related to SARS-CoV-2 (Multimedia Appendix 1), they add to the proof for the public's increasing interest in immunity following the spread of COVID-19.

The interests of the public are always inextricably tied to the media [25]. As such, rather than supplying sufficient information on the subjects that the public are interested in, news outlets have a tendency to instead provide distracting and shocking content [26]. During a pandemic, web-based media outlets focus on reporting related news, which causes worry about the disease spreading rapidly among the public [27]. Such an anxious and worried public has a tendency to search for health information that cannot be trusted [28]. In particular, it has already been reported that incorrect health information, rather than accurate health information, tends to spread on social networks [29], and this can also cause people to adopt improper health measures [30]. Moreover, despite there being information that is important to understand, during pandemics and similar emergency situations, it is likely that the public will miss certain important information about disease prevention amid the flood of information, and the likelihood of people adopting false beliefs or actions will also rise. The public's interests and fears concerning COVID-19 can lead to various negative actions such as the panic buying of health products and the aggressive exclusion of certain racial groups, which, in the end, has a negative effect on preventing the spread of SARS-CoV-2 [31,32]. In contrast, during the SARS pandemic, it was known that a certain amount of worry and panic had a positive effect on preventative actions [33], which suggested that because of knowledge deficiencies and areas of irrational risk perception, mediation was needed [33]. Moreover, research on a section of the United States showed that in 2009, newspaper reports on the H1N1 flu had a positive effect on people's desire to seek prevention measures.

The internet can be used to disseminate information across the globe, and it supplies individually tailored and interactive information. As such, it is an extremely appropriate platform for promoting public health interventions [34,35]. In fact, it is known that information searches conducted through social networks such as Google, YouTube, and Facebook have a positive effect on gaining knowledge about how to prevent infectious diseases, which results in people taking preventative measures [36]. However, as mentioned earlier, the chances are very high that current internet culture will be applied negatively to the COVID-19 pandemic. Within this context, interventions that can assist people in finding information through search word matching and help with actually preventing and coping with the spread of the disease are a key issue with regard to the expansion of proper health literacy. Although such measures can be considered beyond the scope of public health science, during a disaster situation, health authorities must make active interventions to prevent the spread of a disease and do so from a public health science and health information perspective. It is clear that doctors and health care workers must strive to provide accurate information to the public.

It is a known fact that products such as vitamins are helpful for strengthening one's immune system, and they may have a positive effect on decreasing the rates of infections and deaths caused by SARS-CoV-2 [37,38]. However, although a lack of vitamins can become a medical problem, there is no evidence that an overconsumption of vitamins prevents the transmission of COVID-19. It is known that in seriously ill patients, high doses of vitamin C can be helpful for treatment [39]. However, although it has medical uses, this does not mean that the public should be taking vitamin C to prevent the contraction of COVID-19. The interests of the public, including those of the media, are not necessarily based on accurate medical science. When SARS-CoV-2 emerged, the media disseminated a high volume of reports on immunity and vitamins. Since articles on immunity and vitamins were written mainly about SARS-CoV-2, the public might have believed that vitamin C could prevent SARS-CoV-2 infection. However, although these articles may increase people's knowledge about self-care related to COVID-19, they can also increase incorrect health literacy. There is also a possibility that such articles will be used for commercial purposes. In the initial stages of the COVID-19 outbreak, the public's interest in how to protect themselves from infectious diseases, such as by strengthening one's immune system, grew rapidly. Moreover, an infodemic [40] of misinformation, which spread through various media, became an obstacle to managing public health [41].

During the COVID-19 pandemic, delivering accurate information to the public and correcting false information is the responsibility of experts. Our study confirmed that during the initial phase of the COVID-19 crisis, the public's desire and actions of strengthening their own immune systems were enhanced. Therefore, when considering our results in conjunction with earlier findings, it is clear that the initial stage of the spread of an infectious disease is the period in which methods for strengthening accurate knowledge about a disease become the most effective. Moreover, these results also verify the importance of intervention strategies that are meant to prevent the spread of an infectious disease during the early stages.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), which is funded by the Ministry of Education (NRF-2020R1I1A1A01070465).

Authors' Contributions

JL and MBP designed the study, conducted the literature review and statistical analysis, and wrote the manuscript. YK and JYL reviewed related articles and references and wrote the manuscript. JIS and KHL checked the statistical analysis and collected the data. A Kronbichler, LS, A Koyanagi, LJ, SWC, and RAG checked the manuscript as a whole. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 159 KB - jmir_v23i6e26368_app1.docx]

References

1. Coronavirus disease (COVID-19). World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2021-06-01]
2. Cutler DM, Everett W. Thinking outside the pillbox--medication adherence as a priority for health care reform. N Engl J Med 2010 Apr 29;362(17):1553-1555. [doi: 10.1056/NEJMp1002305] [Medline: 20375400]
3. Hoover R, Sheth P, Burde A. Determining the accuracy of open-access databases for identifying commonly prescribed oral medications. J Am Pharm Assoc (2003) 2016 Jan;56(1):37-40. [doi: 10.1016/j.japh.2015.11.009] [Medline: 26802918]
4. Olshannikova E, Olsson T, Huhtamäki J, Kärkkäinen H. Conceptualizing big social data. J Big Data 2017 Jan 25;4(1):1-19 [FREE Full text] [doi: 10.1186/s40537-017-0063-x]
5. Arora VS, McKee M, Stuckler D. Google Trends: Opportunities and limitations in health and health policy research. Health Policy 2019 Mar;123(3):338-341. [doi: 10.1016/j.healthpol.2019.01.001] [Medline: 30660346]
6. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564. [doi: 10.1086/630200] [Medline: 19845471]
7. Dugas AF, Hsieh YH, Levin SR, Pines JM, Mareiniss DP, Mohareb A, et al. Google Flu Trends: correlation with emergency department influenza rates and crowding metrics. Clin Infect Dis 2012 Feb 15;54(4):463-469 [FREE Full text] [doi: 10.1093/cid/cir883] [Medline: 22230244]
8. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205. [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]
9. Bhattacharya I, Ramachandran A, Bhattacharya J, Dogra NK. Google trends for formulating GIS mapping of disease outbreaks in India. International Journal of Geoinformatics 2013;9(3):9-19 [FREE Full text]
10. Ayers JW, Althouse BM, Allem JP, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. Am J Prev Med 2013 May;44(5):520-525. [doi: 10.1016/j.amepre.2013.01.012] [Medline: 23597817]
11. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
12. Chowell G, Bertozzi SM, Colchero MA, Lopez-Gatell H, Alpuche-Aranda C, Hernandez M, et al. Severe respiratory disease concurrent with the circulation of H1N1 influenza. N Engl J Med 2009 Aug 13;361(7):674-679. [doi: 10.1056/NEJMoa0904023] [Medline: 19564633]
13. Cleaton JM, Viboud C, Simonsen L, Hurtado AM, Chowell G. Characterizing Ebola transmission patterns based on internet news reports. Clin Infect Dis 2016 Jan 01;62(1):24-31 [FREE Full text] [doi: 10.1093/cid/civ748] [Medline: 26338786]
14. Reintjes R, Das E, Klemm C, Richardus JH, Keßler V, Ahmad A. "Pandemic Public Health Paradox": Time series analysis of the 2009/10 Influenza A / H1N1 epidemiology, media attention, risk perception and public reactions in 5 European countries. PLoS One 2016 Mar 16;11(3):e0151258. [doi: 10.1371/journal.pone.0151258] [Medline: 26982071]
15. Capala M. Global search engine market share in the top 15 GDP nations (updated for 2020). Alphametic. URL: https://alphametic.com/global-search-engine-market-share [accessed 2021-06-01]
16. Johnson J. Market share of non-desktop search held by Google in Italy from October 2015 to January 2020. Statista. URL: https://www.statista.com/statistics/623054/google-s-market-share-of-non-desktop-search-in-italy/ [accessed 2021-06-01]
17. Fleming A. How to boost your immune system to avoid colds and coronavirus. The Guardian. 2020 Mar 08. URL: https://www.theguardian.com/lifeandstyle/2020/mar/08/how-to-boost-your-immune-system-to-avoid-colds-and-coronavirus [accessed 2021-06-01]
18. Coronavirus pandemic (COVID-19) - Statistics and research. Our World in Data. URL: https://ourworldindata.org/coronavirus [accessed 2021-06-01]
19. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. Br J Health Psychol 2010 Nov;15(Pt 4):797-824 [FREE Full text] [doi: 10.1348/135910710X485826] [Medline: 20109274]
20. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. Ann Intern Med 2020 May 05;172(9):577-582 [FREE Full text] [doi: 10.7326/M20-0504] [Medline: 32150748]
21. Johnson EJ, Hariharan S. Public health awareness: knowledge, attitude and behaviour of the general public on health risks during the H1N1 influenza pandemic. J Public Health (Oxf) 2017 Jan 24;25:333-337 [FREE Full text] [doi: 10.1007/s10389-017-0790-7]

22. Roy D, Tripathy S, Kar SK, Sharma N, Verma SK, Kaushal V. Study of knowledge, attitude, anxiety & perceived mental healthcare need in Indian population during COVID-19 pandemic. Asian J Psychiatr 2020 Jun;51:102083 [FREE Full text] [doi: 10.1016/j.ajp.2020.102083] [Medline: 32283510]
23. Tian H, Liu Y, Li Y, Wu CH, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020 May 08;368(6491):638-642 [FREE Full text] [doi: 10.1126/science.abb6105] [Medline: 32234804]
24. Tsai J, Wilson M. COVID-19: a potential public health problem for homeless populations. Lancet Public Health 2020 Apr;5(4):e186-e187 [FREE Full text] [doi: 10.1016/S2468-2667(20)30053-0] [Medline: 32171054]
25. Strömbäck J. The media and their use of opinion polls: Reflecting and shaping public opinion. In: Holtz-Bacha C, Strömbäck J, editors. Opinion Polls and the Media. London: Palgrave Macmillan; 2012:1-22.
26. Solík M, Višňovský J. The issue of media reflection on socio-cultural recognition and global change. European Journal of Science and Theology 2017 Dec;13(6):249-260 [FREE Full text]
27. Song J, Song TM, Seo DC, Jin DL, Kim JS. Social big data analysis of information spread and perceived infection risk during the 2015 Middle East respiratory syndrome outbreak in South Korea. Cyberpsychol Behav Soc Netw 2017 Jan;20(1):22-29. [doi: 10.1089/cyber.2016.0126] [Medline: 28051336]
28. Baumgartner SE, Hartmann T. The role of health anxiety in online health information search. Cyberpsychol Behav Soc Netw 2011 Oct;14(10):613-618. [doi: 10.1089/cyber.2010.0425] [Medline: 21548797]
29. 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]
30. Park M, Sun Y, McLaughlin ML. Social media propagation of content promoting risky health behavior. Cyberpsychol Behav Soc Netw 2017 May;20(5):278-285. [doi: 10.1089/cyber.2016.0698] [Medline: 28498046]
31. Devakumar D, Shannon G, Bhopal SS, Abubakar I. Racism and discrimination in COVID-19 responses. Lancet 2020 Apr 11;395(10231):1194 [FREE Full text] [doi: 10.1016/S0140-6736(20)30792-3] [Medline: 32246915]
32. Leung CC, Lam TH, Cheng KK. Mass masking in the COVID-19 epidemic: people need guidance. Lancet 2020 Mar 21;395(10228):945 [FREE Full text] [doi: 10.1016/S0140-6736(20)30520-1] [Medline: 32142626]
33. Leung GM, Lam TH, Ho LM, Ho SY, Chan BHY, Wong IOL, et al. The impact of community psychological responses on outbreak control for severe acute respiratory syndrome in Hong Kong. J Epidemiol Community Health 2003 Nov;57(11):857-863 [FREE Full text] [doi: 10.1136/jech.57.11.857] [Medline: 14600110]
34. Gesualdo F, Romano M, Pandolfi E, Rizzo C, Ravà L, Lucente D, et al. Surfing the web during pandemic flu: availability of World Health Organization recommendations on prevention. BMC Public Health 2010 Sep 20;10:561 [FREE Full text] [doi: 10.1186/1471-2458-10-561] [Medline: 20854690]
35. Fung ICH, Fu KW, Ying Y, Schaible B, Hao Y, Chan CH, et al. Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. Infect Dis Poverty 2013 Dec 20;2(1):31 [FREE Full text] [doi: 10.1186/2049-9957-2-31] [Medline: 24359669]
36. Oh SH, Lee SY, Han C. The effects of social media use on preventive behaviors during infectious disease outbreaks: The mediating role of self-relevant emotions and public risk perception. Health Commun 2021 Jul;36(8):972-981. [doi: 10.1080/10410236.2020.1724639] [Medline: 32064932]
37. Grant WB, Lahore H, McDonnell SL, Baggerly CA, French CB, Aliano JL, et al. Evidence that vitamin D supplementation could reduce risk of influenza and COVID-19 infections and deaths. Nutrients 2020 Apr 02;12(4):988 [FREE Full text] [doi: 10.3390/nu12040988] [Medline: 32252338]
38. Biesalski HK. Vitamin D deficiency and co-morbidities in COVID-19 patients – A fatal relationship? NFS Journal 2020 Aug;20:10-21 [FREE Full text] [doi: 10.1016/j.nfs.2020.06.001]
39. Carr AC. A new clinical trial to test high-dose vitamin C in patients with COVID-19. Crit Care 2020 Apr 07;24(1):133 [FREE Full text] [doi: 10.1186/s13054-020-02851-4] [Medline: 32264963]
40. Hua J, Shaw R. Corona virus (COVID-19) "infodemic" and emerging issues through a data lens: The case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]
41. Mian A, Khan S. Coronavirus: the spread of misinformation. BMC Med 2020 Mar 18;18(1):89 [FREE Full text] [doi: 10.1186/s12916-020-01556-3] [Medline: 32188445]

Abbreviations

CDC: Centers for Disease Control and Prevention
NRF: National Research Foundation of Korea
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by C Basch; submitted 08.12.20; peer-reviewed by D Vollmer Dahlke, T Freeman, S Pesälä; comments to author 14.01.21; revised version received 04.02.21; accepted 15.04.21; published 18.06.21.

Please cite as:
Lee J, Kwan Y, Lee JY, Shin JI, Lee KH, Hong SH, Han YJ, Kronbichler A, Smith L, Koyanagi A, Jacob L, Choi S, Ghayda RA, Park MB
Public Interest in Immunity and the Justification for Intervention in the Early Stages of the COVID-19 Pandemic: Analysis of Google Trends Data
J Med Internet Res 2021;23(6):e26368
URL: https://www.jmir.org/2021/6/e26368
doi:10.2196/26368
PMID:34038375

Original Paper

Influence of Baseline User Characteristics and Early Use Patterns (24-Hour) on Long-Term Adherence and Effectiveness of a Web-Based Weight Loss Randomized Controlled Trial: Latent Profile Analysis

Andre Q Andrade[1], MD, PhD; Alline Beleigoli[2,3,4], MD, PhD; Maria De Fatima Diniz[4], MD, PhD; Antonio Luiz Ribeiro[5,6], MD, PhD

[1]Quality Use of Medicines and Pharmacy Research Centre, University of South Australia, Adelaide, Australia
[2]Flinders Digital Health Research Centre, Flinders University, Adelaide, Australia
[3]Caring Futures Institute, Flinders University, Adelaide, Australia
[4]Adult Health Sciences Post Graduation Course, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[5]Centre of Telehealth of the Hospital das Clinicas da UFMG, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[6]Faculty of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Corresponding Author:
Alline Beleigoli, MD, PhD
Flinders Digital Health Research Centre
Flinders University
Tonsley Boulevard
Tonsley
Adelaide,
Australia
Phone: 61 08 2013303
Email: alline.beleigoli@flinders.edu.au

Abstract

Background: Low adherence to real-world online weight loss interventions reduces long-term efficacy. Baseline characteristics and use patterns are determinants of long-term adherence, but we lack cohesive models to guide how to adapt interventions to users' needs. We also lack information whether very early use patterns (24 hours) help describe users and predict interventions they would benefit from.

Objective: We aim to understand the impact of users' baseline characteristics and early (initial 24 hours) use patterns of a web platform for weight loss on user adherence and weight loss in the long term (24 weeks).

Methods: We analyzed data from the POEmaS randomized controlled trial, a study that compared the effectiveness of a weight loss platform with or without coaching and a control approach. Data included baseline behavior and use logs from the initial 24 hours after platform access. Latent profile analysis (LPA) was used to identify classes, and Kruskal-Wallis was used to test whether class membership was associated with long-term (24 weeks) adherence and weight loss.

Results: Among 828 participants assigned to intervention arms, 3 classes were identified through LPA: class 1 (better baseline health habits and high 24-hour platform use); class 2 (better than average health habits, but low 24-hour platform use); class 3 (worse baseline health habits and low 24-hour platform use). Class membership was associated with long-term adherence ($P<.001$), and class 3 members had the lowest adherence. Weight loss was not associated with class membership ($P=.49$), regardless of the intervention arm (platform only or platform + coach). However, class 2 users assigned to platform + coach lost more weight than those assigned to platform only ($P=.02$).

Conclusions: Baseline questionnaires and use data from the first 24 hours after log-in allowed distinguishing classes, which were associated with long-term adherence. This suggests that this classification might be a useful guide to improve adherence and assign interventions to individual users.

Trial Registration: ClinicalTrials.gov NCT03435445; https://clinicaltrials.gov/ct2/show/NCT03435445

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-018-5882-y

(J Med Internet Res 2021;23(6):e26421) doi:10.2196/26421

KEYWORDS

obesity; overweight; web platform; digital health; engagement; latent profile analysis; online interventions; use data; weight loss; weight loss platform

Introduction

More than a third (36%) of adults worldwide are overweight (BMI >25 kg/m^2) or obese (BMI >30 kg/m^2) [1], which increases the risk for metabolic, cardiovascular, and neoplastic diseases. Daily behaviors such as excessive energy intake and lack of physical exercise are major determinants of excess weight. The advent of the internet and the ubiquity of digital technologies, such as smartphones, has created an opportunity for behavior change at scale.

Interventions delivered using digital technology (digital health interventions) and, in particular, web-based platforms, are promising tools to behavior change and weight management promotion [2]. However, as demonstrated in systematic reviews, their effect is typically short term [3,4]. Technology adherence, defined as the percentage of users who follow an intervention's intended use pattern, affects the implementation of large-scale digital health interventions [5]. In many ways, adherence has been the Achilles heel of any health intervention aimed at behavior change. The law of attrition, which states that a considerable proportion of users will stop using eHealth interventions, is one of the few laws that has remained consistent throughout the years of digital health intervention development [6]. There seems to be some agreement as to how we can design for adherence, using concepts such as tailoring and fit and combining personal support with digital functionalities [7]. However, developers face several challenges to formulating personalized digital health interventions, given the individual variation of patient characteristics and motivation to use digital health technology.

Adherence to technology can be explained by environmental, technological, and support variables, as well as individual user demographics and psychological characteristics [8]. User characteristics present intervention designers with two challenges. The first is that different user segments may require different products [9]. By understanding user needs, digital literacy, and habits, interventions can be tailored to users' needs. Tailoring and personalization [10], which also inherently considers individual characteristics, goals, and previous motivation, have been shown to be more effective than static nonadaptive interventions [2]. The second challenge is the high variability on the adoption, response, and outcomes of digital behavior change interventions across users [11]. With some users benefiting more than others, it appears that preintervention individual characteristics are important determinants of adopting a new behavior [12,13]. Given health apps' low retention rates after the first few days [14], the capacity to adequately adapt apps based on data available on day 1 (baseline and initial use data) could improve user experiences and potentially improve follow-up.

To meet these challenges and improve understanding about user adherence in weight loss interventions delivered through web platforms, we will perform subanalysis of the POEmaS randomized controlled trial (RCT). This provides an opportunity to study users' baseline characteristics and the impact of these characteristics on adherence and behavior change. Our main goals are to understand the impact of users' characteristics and the first 24-hour use patterns of a web platform on user adherence within 24 weeks and evaluate whether preintervention characteristics affect the intervention effect within 24 weeks.

Methods

Study Design

The Online Platform for Healthy Weight Loss (POEmaS, from the abbreviation in Portuguese) study was a 3-arm (1:1:1), parallel, RCT evaluating the efficacy of a coach-supported online platform for promoting weight loss [15]. The protocol [16] and main results [15] are described in detail elsewhere. The RCT was registered at ClinicalTrials.gov (NCT03435445). This is a substudy of the POEmaS RCT that includes data from the two intervention arms only.

Setting

The setting was the university community (all campuses) of the Universidade Federal de Minas Gerais, Brazil.

Recruitment

University students and staff were recruited online using mailing lists and notice boards and through banners and posters across the university campuses reaching current and previous staff, students, and alumni from September to October 2017. They were directed to a website where they received further information about the study, eligibility criteria, and enrollment information.

Eligibility Criteria

Inclusion criteria for the RCT included age 18 to 60 years, BMI ≥25 kg/m^2, intention to lose weight through a behavior change program, and internet access. Exclusion criteria included pregnancy, presence of conditions that demand specific dietary or physical activity recommendations (diabetes, heart failure, coronary artery disease, kidney disease, hepatic disease, cancer, phenylketonuria, celiac disease, food allergies, bariatric surgery history), and participation in any other weight loss program at baseline. For this substudy, participants of the comparison (control group) were excluded since their adherence to the intervention was not measured.

Randomization and Allocation

Those who were eligible were allocated to 1 of 3 study groups using a stratified randomized block design by sex and category of BMI (25 to <30 or ≥30 kg/m^2) using blocks of variable length (either 3 or 6). Participants then received an email with information about the activities available to the group they were allocated to. The random allocation sequence and algorithm for randomization were developed by a team of information technology specialists who did not participate in the recruitment or assessment processes. Those who did not complete the questionnaires about dietary and physical activity habits during the onboarding process could not proceed to the use of the platform.

Intervention

The platform-only group was given access to a weight loss program delivered by a web-based platform. The program was based on diet and physical activity guidelines and on the behavior change wheel [17] model comprising a total of 24 weekly sessions (12 weeks of intensive and 12 weeks of maintenance program). Knowledge/empowerment, goal setting, outcomes expectations, self-monitoring, modeling, social support, personalization, and problem solving were the behavior change techniques applied [18]. They were delivered by a range of software functionalities such as short educational readings and videos, graphical and interactive tools, qualitative and quantitative (food diary) dietary monitoring, physical activity self-monitoring tasks, interactive games that created opportunities to invite friends and adopt healthy habits in daily life, and an online social network moderated by physicians and dietitians (Table 1). Personalized feedback on achievements and suggestions of strategies to improve their success in accordance to their individual goals were provided to participants from the fourth week of the intervention. This personalized feedback was generated by a computational algorithm that took into account the goals set by each individual participant, data on habits reported by them in initial questionnaires and through the self-monitoring tools, and patterns of use of the platform (types of functionalities most and least used by each participant) during the first 4 weeks. This platform was adapted from commercial software that had been used for multiple workforce behavior change/wellness interventions in Brazil and is described at length in Beleigoli et al [16].

The platform + coach group followed the same 24-week weight loss program delivered by the platform enhanced by a 12-week initial course of online personalized education and feedback by a dietitian. In addition to the interactions with the platform depicted in Table 1, participants in this group could interact with a dietitian through a private chat forum embedded in the platform. The interactions could be initiated by either side. There was not a limit around the number of interactions that could occur.

Table 1. Interactions participants could have with the platform in the two intervention groups.

Category	Interaction
Social	Open the online social network; create or interact with a personal or public post or comment
Self-reporting	Report of a behavior or outcome or entries in the food diary
Content	Read a text or watch a video
Profile	Interaction with profile for personal data visualization

First 24 Hours of the Intervention

The first 24 hours of the intervention were dedicated to onboarding, including introduction and basic tutorial. Users were informed about which group they were assigned to during onboarding and by email. Both intervention groups (platform and platform + coaching) had access to the same platform and content. Additionally, the platform + coaching group had access to a prerecorded video of the coach welcoming participants to the group. No additional messages were exchanged during the initial 24 hours.

Data Collection

Baseline data were collected during enrollment and included self-reported anthropometry (weight, height), physical activity and sedentary behavior (short form of the International Physical Activity Questionnaire [19]), health perception (single question derived from the Short Form 36), overall health status (modified from the Short Form Six-Dimension [20]) and exercise stage of change (adapted from Sutton et al [21]); all questionnaires in the validated Brazilian version. We also collected information about health goals, education, and marital status.

To evaluate food content, questions were asked about the total number of daily servings of fruit and vegetables, whole carbohydrates sources, and weekly servings of sweetened beverages and snack foods. This questionnaire was adapted from Molina et al [22]. Eating behaviors were evaluated by including number of days having breakfast, number of daily meals, frequency of take-away food consumption, and eating while watching television. These questions were derived from a dietary guideline tailored to the Brazilian population [23].

For both the food content and eating behavior questionnaires, each question was scored ranging from 1 to 5, with higher scores reflecting higher frequency of recommended habits (eg, consumption of fruits and vegetables and number of days having breakfast) and lower frequency of eating habits that should be avoided (eg, daily consumption of sweetened beverages or eating while watching television). Scores were then averaged to find the final result for each questionnaire.

Additionally, user interactions (clicks and page transitions) were recorded and stored in access logs. Access data were time stamped, categorized according to type of interaction (see Table 1), and filtered to only include data from 24 hours after initial log-in.

Due to intervention design, the initial interaction of control group with the platform was very limited. Additionally, continuous adherence was also limited and, therefore, we excluded this group from the current analysis.

Outcomes

For this substudy, adherence and weight loss within 24 weeks of the beginning of the intervention were the outcomes of interest. Adherence was defined as the number of distinct days each user logged in to the platform during the first 24 weeks of the intervention. Weight was self-reported by participants at baseline and 24 weeks after the beginning of the intervention. Weight loss was defined as the difference in weight between these two points in time.

Statistical Analysis

To find the main categories of weight loss intervention users, we analyzed data using latent profile analysis (LPA), a mixture model suited for recovering hidden groups from data [24]. This method is a commonly used cluster method that categorizes individuals into 1 of n groups (or classes), making the findings easier to interpret while losing the least amount of information.

To select the variables used, we initially included all variables describing baseline characteristics (age, eating, and exercise behaviors) and user interactions with the system in 24 hours. Variables that prevented the model from converging or reduced accuracy were removed. The period considered to collect user interactions (24 hours) was selected due to a high number of users (27%) who didn’t return to the app after the first session.

To prevent disproportionate influence from variables with different magnitudes, continuous data were normalized (mean = 0 and standard deviation = 1). Subsequently, we ran the LPA algorithm. To find the best fit, we compared models using different parameters such as the number of possible classes and which variables should be included in the model. We progressively removed variables from the model and ran the algorithm with a growing number of classes, starting at 1 and moving up to 6 classes. The best model was selected using two criteria: Bayesian information criterion (BIC) and highest entropy. The BIC has been shown as a good indicator to select the correct number of classes in mixture models [25]. A lower BIC means a better fit. Entropy is a measure of accuracy (ie, how much each individual belongs to the class it was assigned by the algorithm). High entropy means the individuals were well classified in the groups. Models with similar BIC and entropy were selected based on theoretical plausibility and parsimony.

Finally, we analyzed how classification into the newfound groups predicted future adherence to the intervention and weight loss at 24 weeks (6 months). To test whether the groups are related to long-term adherence and weight loss, we performed the Kruskal-Wallis nonparametric test with a .05 significance level.

Missing weight data at 24 weeks was imputed by fitting logistic and linear regression models with both the predictors and outcome as well as with other variables regarded as important to explain the missing values. This procedure generated 5 complete data sets, which were used to estimate the association between group allocation and primary and secondary outcomes.

We used Python packages Pandas [26] and Scipy [27] for data preprocessing and statistical analysis, respectively. For the LPA, we used the R package tidyLPA [28]. Multiple imputation was performed using SPSS (version 18, IBM Corp).

Ethics

The study was approved by the ethics committee of the Universidade Federal de Minas Gerais (CAAE: 73545717.5.0000.5149). All participants signed an online informed consent form prior to recruitment.

Results

Of the 1298 patients enrolled in the POEmaS trial, 828 were randomly assigned to the intervention arms (420 in the platform only group and 408 in the platform + coach group) and included in this analysis. Baseline characteristics can be found in Table 2.

The model with best fit was found using the number of self-report interactions during first 24 hours and scores for 4 questionnaires: the Modified Short Form Six-Dimension, Self-health evaluation (extracted from a single question derived from the Short Form 36), Eating behavior (higher scores means higher frequency of recommended eating habits), and food content (higher scores means frequent consumption of recommended foods such as fruits and vegetables). The optimal number of classes according to the BIC measure was 3, with a BIC 14,833 (see Figure 1).

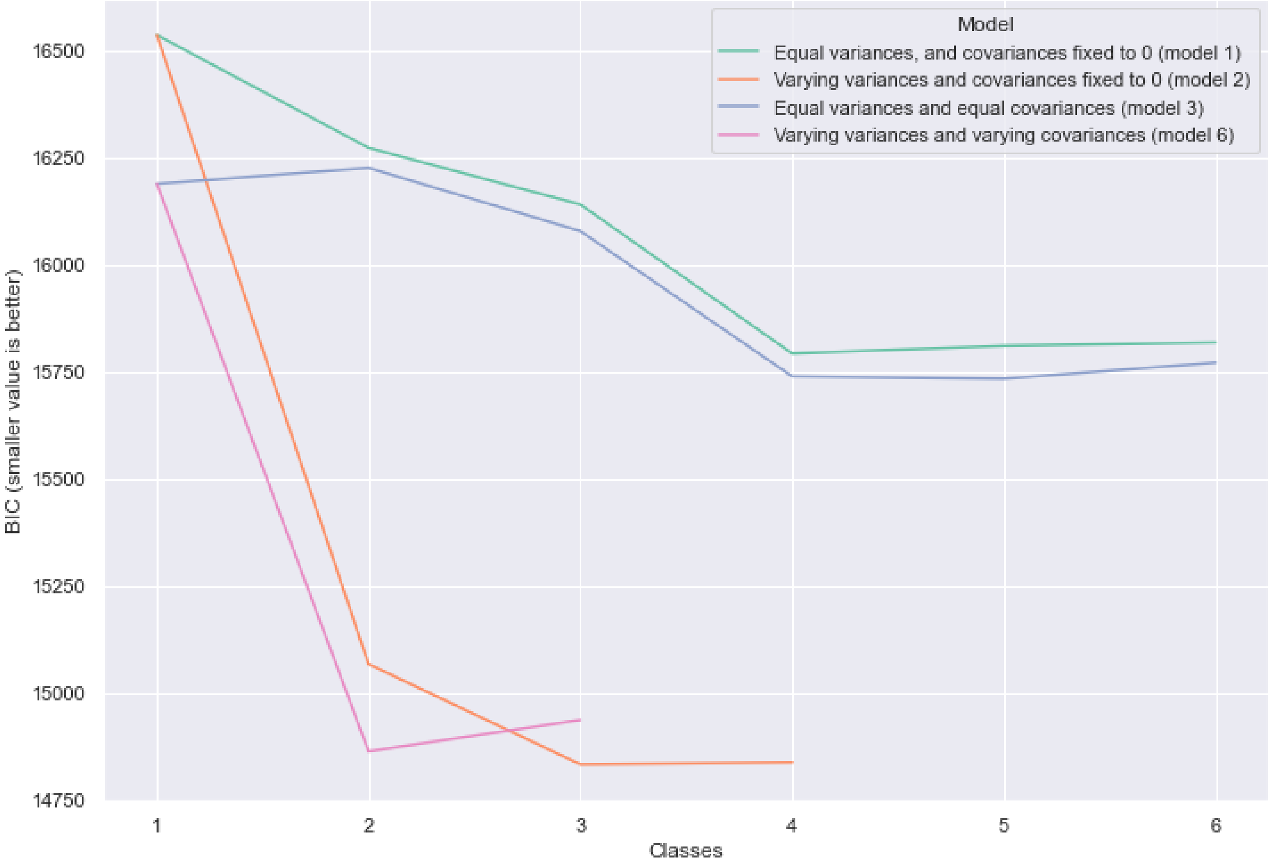
Table 2. Baseline characteristics of the groups included in the analysis[a].

Characteristic	Platform only group (n=420)	Platform + coach group (n=408)
Weight (kg), mean (95% CI)	83.4 (81.7, 85.0)	82.3 (80.8, 83.7)
Age (years), mean (95% CI)	34.4 (33.4, 35.6)	33.0 (31.9, 34.0)
BMI (kg/m^2), mean (95% CI)	30.12 (29.67, 30.58)	29.85 (29.44, 30.26)
Female, n (%)	315 (75.0)	319 (78.2)
Stages of change for physical activity[b], n (%)		
Precontemplation	12 (2.9)	15 (3.7)
Contemplation	163 (38.8)	126 (30.9)
Preparation	96 (22.9)	94 (23.0)
Action	79 (18.8)	91 (22.3)
Maintenance	45 (10.7)	43 (10.5)
Did not respond	25 (6.0)	39 (9.6)

[a] Control group participants were excluded from this analysis since they had limited interaction with the platform.

[b] Precontemplation: not intending to engage in physical activity within 6 months; contemplation: intending to engage in physical activity within 6 months; preparation: intending to engage in physical activity within 30 days; action: physically active for less than 6 months; maintenance: physically active for more than 6 months.

Figure 1. Comparison of different numbers of classes and models as measured by the Bayesian information criterion indicator.



The 3 distinct groups identified presented significant differences between them (see Figure 2), and the accuracy of the model in correctly classifying individuals was 0.89 as measured by the entropy indicator (entropy for the 2-class model was 0.82 and for the 4-class model was 0.79). Their characteristics are described in Table 3.

There were no significant differences in terms of marital status. Individuals in class 1 had more education years than classes 2 and 3, with 50% graduate students or alumni compared to 40% and 20% in classes 2 and 3, respectively (*P*<.001).

Individuals in class 1 were more likely to select a goal than those in classes 2 and 3, but the choice of goal (lose weight or improve habits or health) was similar in all classes.

At baseline, the classes were significantly different in terms of BMI (*P*=.04), although the difference was small and not clinically important. Users classified as class 1 had a lower BMI, while users in class 3 had a higher BMI than the other classes. Classes also differed according to the number of users assigned to different interventions, with 56% (107/190) of users in class 3 assigned to coaching as compared to 41% (107/260) in class 1.

Figure 2. Classes identified in the latent profile analysis as described by the normalized variables.

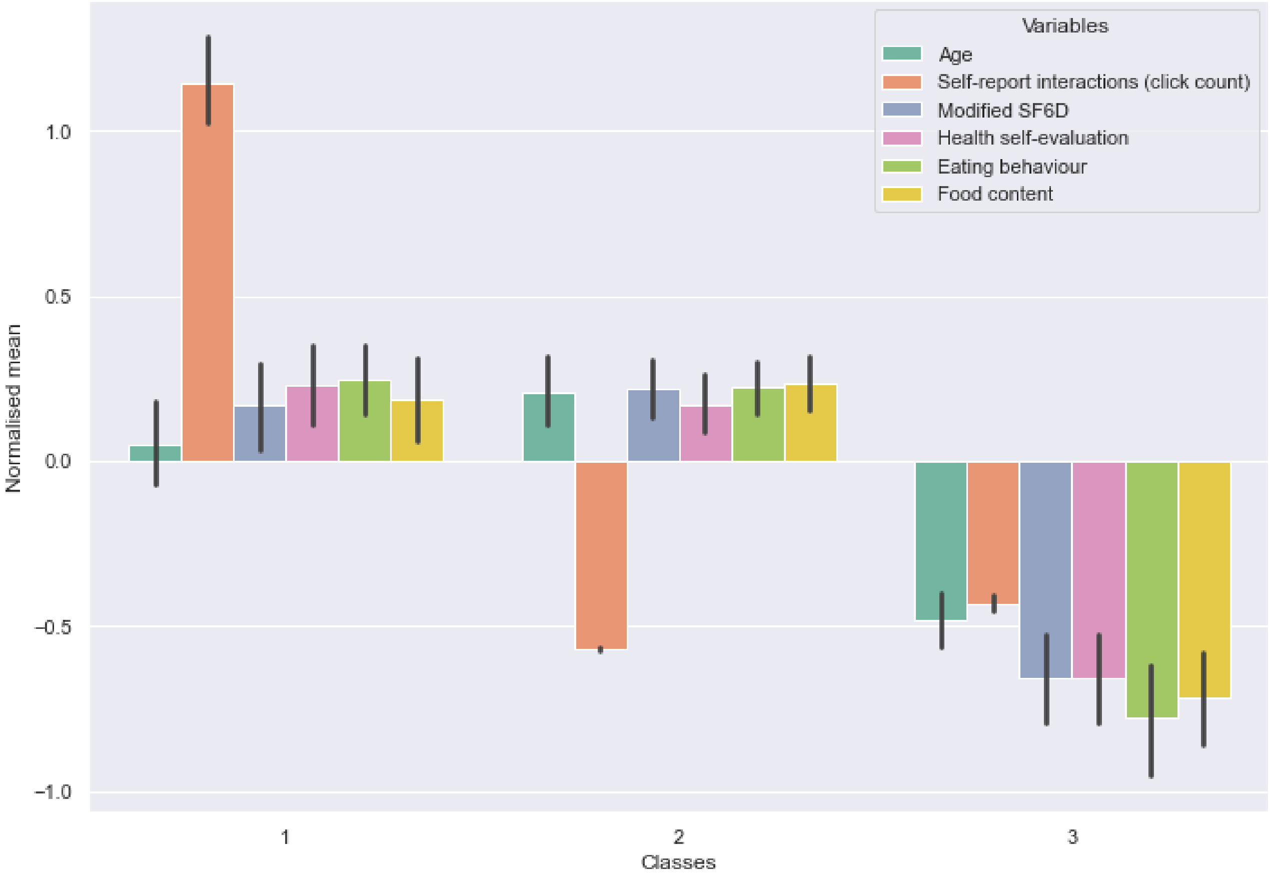


Table 3. Characteristics of users according to latent profile groupings.

Covariates	Class 1 (n=260)	Class 2 (n=378)	Class 3 (n=190)	*P* value
Characteristics used to determine classes, mean (SD)				
Age (years)	34 (11)	36 (11)	28 (6)	—[a]
Food content score	28.7 (4.89)	28.9 (4.22)	24.2 (4.67)	—
Eating behavior score	12.5 (1.71)	12.4 (1.59)	10.4 (2.32)	—
Self-report interactions click count	12.3 (7.77)	0.2 (0.44)	1.2 (1.26)	—
Health self-evaluation	0.64 (0.15)	0.63 (0.14)	0.5 (0.15)	—
Modified SF6D[b]	0.73 (0.14)	0.74 (0.12)	0.62 (0.13)	—
Other characteristics at baseline				
Female, n (%)	200 (76.9)	294 (77.8)	140 (73.7)	.55
BMI, mean (SD)	29.46 (4.15)	30.07 (4.63)	30.54 (4.65)	.04
Assigned to coaching, n (%)	107 (41.2)	194 (51.3)	107 (56.3)	<.001
Stage of change (physical activity) at baseline, % (95% CI)				<.001
Precontemplation	3.9 (1.5, 6.2)	2.4 (0.8, 3.9)	4.2 (1.4, 7.1)	—
Contemplation	20.0 (15.1, 24.9)	19.8 (15.8, 23.9)	33.2 (26.5, 39.9)	—
Preparation	30.8 (25.2, 36.4)	33.1 (28.3, 37.8)	44.2 (37.2, 51.3)	—
Action	12.7 (8.7, 16.7)	10.3 (7.3, 13.4)	8.4 (4.5, 12.4)	—
Maintenance	29.6 (24.1, 35.2)	20.4 (16.3, 24.4)	8.4 (4.5, 12.4)	—
Did not respond	3.1 (1.0, 5.2)	14.0 (10.5, 17.5)	1.6 (0, 3.4)	—
Marital status, % (95% CI)				.27
Did not respond	38.9 (32.9, 44.8)	64.6 (59.7, 69.4)	58.4 (51.4, 65.4)	—
Married	23.1 (18.0, 28.2)	12.7 (9.3, 16.1)	10.5 (6.2, 14.9)	—
Divorced	3.9 (1.5, 6.2)	2.7 (1.0, 4.3)	1.58 (0, 3.4)	—
Single	33.9 (28.1, 39.6)	19.6 (15.6, 23.6)	29.5 (23.0, 36.0)	—
Widowed	0.4 (0, 1.1)	0.5 (0, 1.26)	0 (0, 0)	—
Health goal, % (95% CI)				<.001
Did not respond	41.5 (35.6, 47.5)	67.5 (62.7, 72.2)	60.5 (53.6, 67.5)	—
Lose weight	44.2 (38.2, 50.3)	24.1 (19.8, 28.4)	30.5 (24.0, 37.1)	—
Improve habits or health	14.2 (10.0, 18.4)	8.5 (5.7, 11.3)	9.0 (4.9, 13.0)	—
Highest educational degree, % (95% CI)				<.001
Did not respond	39.6 (33.7, 45.6)	65.1 (60.3, 70.0)	57.9 (50.9, 64.9)	—
High school	9.2 (5.7, 12.8)	4.76 (2.6, 6.9)	12.1 (7.5, 16.7)	—
University degree	20.8 (15.8, 25.7)	16.1 (12.4, 19.9)	21.6 (15.7, 27.4)	—
Postgraduate degree	30.4 (24.8, 36.0)	14.0 (10.5, 17.5)	8.4 (4.5, 12.4)	—

[a] Not available.

[b] SF6D: Short Form Six-Dimension.

Associations With Weight Loss and Program Adherence at 24 Weeks

There was a significant difference ($P<.001$) among the classes regarding adherence. Class 1 users were more adherent to the intervention than other users, followed by users in class 2. There was no statistically significant difference in weight loss at 24 weeks (Table 4).

To control for the potential influence of intervention assignment at baseline, we stratified the classes by intervention. The stratification did not change the results. The number of sessions was still significantly different within the platform only group and platform + coach group. Weight loss was not significantly different among users in different classes.

Finally, to investigate whether any particular class of individuals would benefit from different interventions, we stratified participants by class and analyzed their weight change at 24 weeks (see Table 5). Individuals in class 2 who were assigned to the platform + coach intervention had a significantly larger weight loss than those assigned to platform only. The other classes showed no significant difference.

Table 4. Number of sessions and weight change at 24 weeks by class.

Outcome	Class 1	Class 2	Class 3	*P* value
Sessions completed in 24 weeks, mean (SD)	9 (19)	6 (12)	4 (6)	<.001
Weight change at 24 weeks (kg), change (SD)	–1.3 (3.6)	–1.2 (3.7)	–1.5 (3.2)	.49
Platform only				
Sessions completed in 24 weeks, mean (SD)	7 (17)	5 (7)	4 (6)	.001
Weight change at 24 weeks (kg), change (SD)	–1.2 (3.5)	–0.8 (3.4)	–1.4 (3.3)	.27
Platform + coach				
Sessions completed in 24 weeks, mean (SD)	12 (22)	7 (15)	4 (5)	<.001
Weight change at 24 weeks (kg), change (SD)	–1.3 (3.7)	–1.7 (3.8)	–1.6 (3.2)	.12

Table 5. Differences in weight change at 24 weeks by class.

Outcome	Platform only	Platform + coach	*P* value
Class 1			
Number (%)	153 (59)	107 (41)	—[a]
Mean weight loss in 24 weeks (kg), mean (SD)	–1.24 (3.48)	–1.34 (3.70)	.93
Class 2			
Number (%)	184 (49)	194 (51)	—
Mean weight loss in 24 weeks (kg), mean (SD)	–0.80 (3.43)	–1.67 (3.82)	.02
Class 3			
Number (%)	83 (44)	107 (56)	—
Mean weight loss in 24 weeks (kg), mean (SD)	–1.41 (3.32)	–1.61 (3.18)	.76

[a] Not available.

Discussion

Principal Findings

Our data suggest that 3 class groups of users that differ by preintervention characteristics and use within the first 24 hours of the intervention are related to adherence and weight loss outcomes in an RCT testing a web-based platform with and without online coaching for adults with overweight or obesity. The first group (class 1) was composed of users with better eating habits and higher use in the first 24 hours with a lower baseline BMI. The second and largest group (class 2) was balanced in terms of both healthy/unhealthy habits and low/high first 24-hour use. This group was slightly older and had an average BMI similar to that of the whole population. The third and smallest group (class 3) was formed by users with the worst habits and lowest use in 24 hours. This class had higher baseline BMI and younger age than the other groups.

Analyzing the groups' characteristics at baseline and their longitudinal behavior reveals some insights to help plan future interventions. The first insight is that clustering using 24-hour and baseline data was predictive of higher adherence to the platform 24 weekly sessions. Early identification of the class 2 individuals can help intervention developers adopt different strategies to promote adherence, such as active communication or change in platform content/value proposition. This suggestion is further reinforced by the differences in weight loss between class 2 individuals assigned to platform and platform + coach. The coaching introduces an active component to the intervention that seems better suited to these individuals and resulted in improved weight loss in this study. Enhancing motivation is a key coaching role, so it probably removes some of the effect of preintervention motivation [29].

The second and most important finding, but perhaps least surprising, was that adherence was highest among the class 1 individuals and smallest in class 3. This suggests that this digital health intervention was used by individuals who needed it the least and was dropped by those who would benefit most from it. The path to changing behaviors associated with weight loss depends on a combination of emotional factors, motivation, knowledge, and external factors. Behavior change theories and frameworks, such as the transtheoretical model [13] and the behavior change wheel [17], posit we should consider current individual needs and stages of change to design successful strategies. This is reflected by the higher proportion of users in the action or maintenance phase (physical exercise) in class 1 compared to class 3 individuals. It is also reflected in the larger proportion of individuals choosing health goals, a characteristic of individuals in later phases of change.

This finding reinforces the opportunity for tailoring apps to user current needs. Commonly offered functionalities, such as calorie counters, are better suited to individuals in preparation and action phase. This may explain why studies show a higher baseline BMI and poorer health status in noncompleters [30]. Results also suggest goal setting as a desired functionality for class 1 individuals. This is recommended, given the evidence that goal setting promotes behavior change [31,32]. The 3 groups identified in this study provide an empirical starting point to guide development of different products for different segments.

An innovation in this study was to use the initial 24 hours of user interactions as a latent profile determinant. These interactions may carry information on digital literacy, prior motivation, and whether users immediately perceive the interventions' value. Early intervention use patterns will vary according to onboarding process and interaction flow. Therefore, it is envisaged that any system collecting interaction data to group users into the 3 classes will require identifying which interaction is most reflective of the initial variable engagement. Finally, in real-world interventions, early interaction may be influenced by factors unrelated to motivation, such as device (mobile vs desktop) or environment where enrollment occurs (work vs home). When known, these factor should be considered to calibrate interaction information.

In this study, the type of intervention had a small but significant influence on the latent profile classification, evidenced by the higher proportion of platform only users in class 1. This was unforeseen, given that data used to discriminate the classes were limited to that collected during the initial 24 hours of the intervention. Additionally, users assigned to different interventions (platform only and platform + coaching) received remarkably similar interventions, with two differences: users in the second group were informed they were assigned to coaching, and they saw a prerecorded video with a welcoming message by the coach. However, our analysis showed that intervention assignment was not the main driver of adherence predicted by the latent profile. Segmented analysis of the platform only and platform + coaching users revealed similar results of differences in adherence between classes.

Strengths and Limitations

Our understanding about users of digital health interventions aimed at promoting weight loss was facilitated by the very inclusive and pragmatic strategy for recruiting participants. The strategy used for recruiting and the low barrier for joining the intervention mimicked real-life interventions, allowing us to measure how users react outside of controlled trial settings. The use of LPA as a cluster method allowed an improved understanding of users of weight loss interventions. Furthermore, due to the longitudinal nature and multiple repeated measures of outcomes collected, we could analyze the power of the identified classes at baseline in predicting weight loss and adherence at 24 weeks.

Limitations of this study include the self-report nature of questionnaire data, which can carry misclassification biases. Given recruitment happened at a university, the study population may not reflect the general population. This can affect the number of classes identified. Finally, enrollment was an active process that may have selected individuals more likely to benefit from it.

Implications to Future Research

The intervention planned for the POEmaS project was designed with an active user in mind. In other words, it required users to log in to the platform, report their activities, and interact in the social network. Such interaction was not suitable for the class 3 individuals, evidenced by the large number of users who didn't return to the intervention. Being able to predict nonadherence allows developers to explore tailoring the intervention, possibly improving intervention effect. The question of how it could be tailored requires further investigation. This study suggests human intervention (coaching) can have a positive effect for class 2 individuals. Additional avenues of investigation include changing the number of active interventions (notifications, emails, or text messages), the content of the messages (focused on contemplation), and understanding the use of gamification techniques aimed at different groups.

This study also highlights the importance of considering the heterogeneity of users when creating and reporting digital health interventions. Any given intervention will affect individuals in different ways based on their baseline habits and motivation. Enrollment strategies may bias the sample (eg, by requiring more active effort before enrollment vs paying participants to complete a study). Setting (clinic, community), barriers to join the intervention (upfront costs, long registration forms, and commitment mechanics), and recruitment method (mailing, targeted communication) all have an impact on selecting the population and, therefore, can influence results. While an additional challenge for creating quality evidence for digital health interventions, careful description of the baseline characteristics and prior motivation of the population groups will create deeper understanding about which intervention is useful for each user.

Conclusions

Three major groups of weight loss intervention users were identified in a large RCT. Baseline questionnaires and use data from the first 24 hours after log in allowed distinguishing classes, which were highly related to long-term adherence. Individuals classified as class 2 lost more weight when assigned to platform + coach than platform only, suggesting this early classification may be a useful guide to intervention selection. These results encourage efforts for early identification of effect predictors to trigger more effective interventions.

Acknowledgments

We acknowledge the priceless contributions of Dr Roberta Alvares, Leonardo Andrade, and Tiago Sizenando to this study. This study was supported with grants from the National Institute of Science and Technology for Health Technology Assessment (CNPq) of Brazil. AB was supported by Prêmio Coordenação de Aperfeiçoamento de Pessoal de Nível Superior de Teses. ALR was supported by CNPq (research fellowship number 310679/2016-8) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (PPM-00428-17). The sponsors did not participate in the study design; data collection, analysis, and interpretation; or the preparation or submission of this report.

Conflicts of Interest

AQA was the owner of the software company that provided the web platform by the time the study was conducted. All other authors declare no conflicts of interest.

References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781. [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
2. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J Med Internet Res 2017 Jun 26;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]
3. 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]
4. Kohl LFM, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
5. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
6. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
7. 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]
8. Ryan C, Bergin M, Wells JS. Theoretical perspectives of adherence to web-based interventions: a scoping review. Int J Behav Med 2017 Jul 20;25(1):17-29. [doi: 10.1007/s12529-017-9678-8] [Medline: 28730402]
9. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Information Soc 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550] [Medline: 26281194]
10. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
11. Tanaka R, Nolan RP. Psychobehavioral profiles to assist tailoring of interventions for patients with hypertension: latent profile analysis. J Med Internet Res 2018 May 11;20(5):e149 [FREE Full text] [doi: 10.2196/jmir.8757] [Medline: 29752248]
12. Bully P, Sánchez A, Zabaleta-del-Olmo E, Pombo H, Grandes G. Evidence from interventions based on theoretical models for lifestyle modification (physical activity, diet, alcohol and tobacco use) in primary care settings: a systematic review. Prev Med 2015 Jul;76 Suppl:S76-S93. [doi: 10.1016/j.ypmed.2014.12.020] [Medline: 25572619]
13. 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]
14. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
15. Beleigoli A, Andrade AQ, Diniz MDF, Ribeiro AL. Personalized web-based weight loss behavior change program with and without dietitian online coaching for adults with overweight and obesity: randomized controlled trial. J Med Internet Res 2020 Nov 05;22(11):e17494 [FREE Full text] [doi: 10.2196/17494] [Medline: 33151151]
16. Beleigoli AM, Queiroz de Andrade A, Haueisen Diniz MDF, Alvares RS, Ribeiro AL. Online platform for healthy weight loss in adults with overweight and obesity—the "POEmaS" project: a randomized controlled trial. BMC Public Health 2018 Aug 01;18(1):945 [FREE Full text] [doi: 10.1186/s12889-018-5882-y] [Medline: 30068393]
17. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

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;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
19. Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ). 2005 Nov. URL: https://sites.google.com/site/theipaq/scoring-protocol/scoring_protocol.pdf?attredirects=0&d=1 [accessed 2021-05-25]
20. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002 Mar;21(2):271-292. [doi: 10.1016/s0167-6296(01)00130-8] [Medline: 11939242]
21. Sutton K, Logue E, Jarjoura D, Baughman K, Smucker W, Capers C. Assessing dietary and exercise stage of change to optimize weight loss interventions. Obes Res 2003 May;11(5):641-652 [FREE Full text] [doi: 10.1038/oby.2003.92] [Medline: 12740454]
22. Molina MDC, Benseñor IM, Cardoso LDO, Velasquez-Melendez G, Drehmer M, Pereira TSS, et al. [Reproducibility and relative validity of the Food Frequency Questionnaire used in the ELSA–Brasil]. Cad Saude Publica 2013 Feb;29(2):379-389 [FREE Full text] [Medline: 23459823]
23. Dietary Guidelines for the Brazilian Population. Brasilia: Primary Health Care Department, Ministry of Health of Brazil; 2015. URL: http://bvsms.saude.gov.br/bvs/publicacoes/dietary_guidelines_brazilian_population.pdf [accessed 2021-05-24]
24. Oberski D. Mixture models: latent profile and latent class analysis. In: Robertson J, Kaptein M, editors. Modern Statistical Methods for HCI. Cham: Springer International Publishing; 2016:275-287.
25. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a monte carlo simulation study. Structural Eq Mod Multidiscip J 2007 Dec 05;14(4):535-569. [doi: 10.1080/10705510701575396]
26. McKinney W. Data structures for statistical computing in Python. Proc 9th Python Sci Conf 2010:1. [doi: 10.25080/Majora-92bf1922-00a]
27. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar 3;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
28. Rosenberg J, Schmidt J, Beymer P, Steingut R. Interface to mclust to easily carry out latent profile analysis (statistical software for R). 2018. URL: https://github.com/jrosen48/tidyLPA [accessed 2021-05-24]
29. Rutten GM, Meis JJM, Hendriks MRC, Hamers FJM, Veenhof C, Kremers SPJ. The contribution of lifestyle coaching of overweight patients in primary care to more autonomous motivation for physical activity and healthy dietary behaviour: results of a longitudinal study. Int J Behav Nutr Phys Act 2014 Jul 16;11:86 [FREE Full text] [doi: 10.1186/s12966-014-0086-z] [Medline: 25027848]
30. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
31. Epton T, Currie S, Armitage CJ. Unique effects of setting goals on behavior change: systematic review and meta-analysis. J Consult Clin Psychol 2017 Dec;85(12):1182-1198. [doi: 10.1037/ccp0000260] [Medline: 29189034]
32. Strecher VJ, Seijts GH, Kok GJ, Latham GP, Glasgow R, DeVellis B, et al. Goal setting as a strategy for health behavior change. Health Educ Q 1995 May;22(2):190-200. [doi: 10.1177/109019819502200207] [Medline: 7622387]

Abbreviations

BIC: Bayesian information criterion
LPA: latent profile analysis
POEmaS: Online Platform for Healthy Weight Loss (acronym in Portuguese)
RCT: randomized controlled trial

Edited by R Kukafka; submitted 11.12.20; peer-reviewed by A Lin, R Asbjornsen; comments to author 16.01.21; revised version received 22.02.21; accepted 06.05.21; published 03.06.21.

Please cite as:
Andrade AQ, Beleigoli A, Diniz MDF, Ribeiro AL
Influence of Baseline User Characteristics and Early Use Patterns (24-Hour) on Long-Term Adherence and Effectiveness of a Web-Based Weight Loss Randomized Controlled Trial: Latent Profile Analysis
J Med Internet Res 2021;23(6):e26421
URL: https://www.jmir.org/2021/6/e26421
doi:10.2196/26421
PMID:34081012

Original Paper

Data Missing Not at Random in Mobile Health Research: Assessment of the Problem and a Case for Sensitivity Analyses

Simon B Goldberg[1,2], PhD; Daniel M Bolt[3], PhD; Richard J Davidson[2,4,5], PhD

[1]Department of Counseling Psychology, University of Wisconsin - Madison, Madison, WI, United States
[2]Center for Healthy Minds, University of Wisconsin - Madison, Madison, WI, United States
[3]Department of Educational Psychology, University of Wisconsin - Madison, Madison, WI, United States
[4]Department of Psychology, University of Wisconsin - Madison, Madison, WI, United States
[5]Department of Psychiatry, University of Wisconsin - Madison, Madison, WI, United States

Corresponding Author:
Simon B Goldberg, PhD
Department of Counseling Psychology
University of Wisconsin - Madison
335 Education Building, 1000 Bascom Mall
University of Wisconsin-Madison
Madison, WI, 53706
United States
Phone: 1 6082658986
Email: sbgoldberg@wisc.edu

Abstract

Background: Missing data are common in mobile health (mHealth) research. There has been little systematic investigation of how missingness is handled statistically in mHealth randomized controlled trials (RCTs). Although some missing data patterns (ie, missing at random [MAR]) may be adequately addressed using modern missing data methods such as multiple imputation and maximum likelihood techniques, these methods do not address bias when data are missing not at random (MNAR). It is typically not possible to determine whether the missing data are MAR. However, higher attrition in active (ie, intervention) versus passive (ie, waitlist or no treatment) conditions in mHealth RCTs raise a strong likelihood of MNAR, such as if active participants who benefit less from the intervention are more likely to drop out.

Objective: This study aims to systematically evaluate differential attrition and methods used for handling missingness in a sample of mHealth RCTs comparing active and passive control conditions. We also aim to illustrate a modern model-based sensitivity analysis and a simpler fixed-value replacement approach that can be used to evaluate the influence of MNAR.

Methods: We reanalyzed attrition rates and predictors of differential attrition in a sample of 36 mHealth RCTs drawn from a recent meta-analysis of smartphone-based mental health interventions. We systematically evaluated the design features related to missingness and its handling. Data from a recent mHealth RCT were used to illustrate 2 sensitivity analysis approaches (pattern-mixture model and fixed-value replacement approach).

Results: Attrition in active conditions was, on average, roughly twice that of passive controls. Differential attrition was higher in larger studies and was associated with the use of MAR-based multiple imputation or maximum likelihood methods. Half of the studies (18/36, 50%) used these modern missing data techniques. None of the 36 mHealth RCTs reviewed conducted a sensitivity analysis to evaluate the possible consequences of data MNAR. A pattern-mixture model and fixed-value replacement sensitivity analysis approaches were introduced. Results from a recent mHealth RCT were shown to be robust to missing data, reflecting worse outcomes in missing versus nonmissing scores in some but not all scenarios. A review of such scenarios helps to qualify the observations of significant treatment effects.

Conclusions: MNAR data because of differential attrition are likely in mHealth RCTs using passive controls. Sensitivity analyses are recommended to allow researchers to assess the potential impact of MNAR on trial results.

(J Med Internet Res 2021;23(6):e26749) doi:10.2196/26749

KEYWORDS

missing data; randomized controlled trial; differential attrition; sensitivity analysis; statistical methodology; mobile phone

Introduction

Background

In the world of mobile health (mHealth), high and rapid attrition is the rule rather than the exception [1]. This *law of attrition* [1] applies both to the use of mHealth interventions in naturalistic settings (eg, internet-based interventions for anxiety and depression [2]) as well as to studies designed to test the efficacy of mHealth interventions (eg, randomized controlled trials [RCTs] of smartphone-based interventions for mental health problems [3]). Attrition in naturalistic settings involves nonusage or discontinuation of usage, whereas attrition in research settings can involve these usage patterns along with dropouts from the study itself [4]. Nonusage and discontinuation of usage can limit the therapeutic potential of mHealth, and the development of methods to increase the sustained uptake of mHealth interventions is an area of active research [5,6]. In research contexts, attrition can not only attenuate therapeutic effects but can also produce additional problems such as reduced statistical power and the introduction of bias. This bias can skew results and limit the generalizability of the study findings (eg, only able to generalize to those who continue use). Various methods have been proposed to prevent attrition in mHealth research (eg, making interventions more engaging, implementing a *run-in* period before randomization, and including remainders and financial incentives [3,7]). However, to date, high attrition appears to be the rule rather than the exception of mHealth research [3].

Attrition in research contexts typically results in missing data. Some exceptions to this may include measures that continue to be assessed regardless of ongoing study participation (eg, smartphone app usage). Nonetheless, decades of methodological work have focused on characterizing the various types of missing data and developing statistical approaches for handling missingness [8,9] (for a more thorough discussion of the various types of missing data and methods for handling them, interested readers are directed to Enders [8] and Graham [9]; for a tutorial specifically geared to nonstatistician mHealth researchers, refer to Blankers et al [10]). There are three basic types of missing data that can be distinguished by their presumed cause as well as their impact on statistical tests [9]. The first and most benign type is data that are missing completely at random (MCAR). For example, an RCT testing a smartphone-based intervention for depression compared with a waitlist control group. In this context, it is common for posttreatment depression scores to be missing for a subset of participants [11]. If the missing data are MCAR, the cases with missing values can be viewed as a random sample of all cases. As such, the missing values did not systematically differ from the observed values. Statistical tests that ignore missing cases (eg, listwise deletion) provide unbiased estimates of parameter values, albeit with reduced statistical power. The second missing data type is data that are missing at random (MAR). For MAR data, the missing value (eg, posttest depression scores) does not depend on the missing value itself (ie, whether the missing score, if observed, would have been high or low) but depends on the observed data. For example, missingness may be more likely for those who had higher depression scores at baseline or were younger, but conditional upon such variables, the outcomes for missing cases resemble those of the observed cases (MCAR is actually a special case of MAR, specifically one in which the missing values are neither associated with observed values or missing values). Similar to MCAR, MAR data can also be analyzed in ways that produce unbiased parameter estimates, provided observed variables on which the missing value depends are included in the imputation and analysis model. Most recent advances in missing data analysis operate under the assumption of MAR. Multiple imputation (MI) and maximum likelihood (ML) are two widely used modern statistical methods that effectively use observed data for unbiased and statistically efficient (ie, not underpowered) analysis of MAR data and can also be applied to MCAR data.

Missing data that are MCAR and MAR are relatively straightforward to handle (so-called *ignorable* missing data [9]). In contrast, data that are missing not at random (MNAR) are a larger problem (ie, not ignorable), particularly when there is a substantial amount of missing data (eg, >5% [9]). For MNAR data, missingness depends on the value of unobserved data. For example, those who would have reported higher depression symptoms posttreatment may be more likely to drop out. The consequences of MNAR in RCTs can be substantial. In our depression RCT example, we found that participants in the active condition (ie, the intervention arm) were more likely to have missing posttest depression scores than the waitlist control (ie, passive control). It is generally impossible to demonstrate that data are MNAR, and simply having differential attrition does not necessarily indicate MNAR data or lead to biased results [12]. Although MNAR cannot be assessed directly, we might speculate that participants in the active condition who did not benefit as much from the smartphone-based intervention may be more likely to drop out of the study (refer to Crutzen et al [13] for similar possibilities offered to explain higher attrition in treatment vs control conditions in health behavior change interventions). This could be because of, for example, the greater effort required to participate in the treatment arm, especially if experiencing higher levels of depression. Such a relationship may well supersede what can be explained by the other measured variables for such subjects. In this case, the likelihood of having an unobserved posttest depression score is dependent on the value of the score itself, had it been observed. Thus, we are under the condition of MNAR. Further, the consequences of the MNAR on the estimation of treatment effects may be substantial, leading to an overestimation of the effect of the treatment because of missing observations. It is theoretically possible that the influence of MNAR data is reversed, with those dropping out experiencing better rather than worse outcomes (eg, dropping out of the study because one's symptoms have already improved). This possibility is viewed as unlikely in related disciplines (eg, addiction research [14]). Lacking data or a strong rationale suggesting that missingness because of improved outcomes is likely in mHealth research, we focus on the more plausible MNAR mechanism of individuals who fail to respond to be those most likely to discontinue study participation.

Unlike MCAR and MAR missingness patterns, MNAR cannot be easily handled in a confident manner. Moreover, MNAR can have multiple causes, making it difficult to develop a single method that can universally address it, even in a single study. As a result, some form of sensitivity analysis is recommended to understand the possible effects of MNAR [12,15]. Advanced tools for evaluating the consequences of MNAR data have been developed, most notably selection models [16] and pattern-mixture models [17]. These models accommodate the joint distribution between the probability of missingness and the observed data and can be powerful techniques for evaluating the impact of MNAR. At once, understanding and implementing selection models or pattern-mixture models is a high bar for many applied researchers who may be faced with MNAR data. These models also involve untestable assumptions whose violations can produce biased results [8,18]. Thus, the application of MNAR procedures is undertaken more in the spirit of understanding the possible implications of missingness rather than explicitly correcting it. This approach is consistent with viewing missing data on a continuum from MAR to MNAR and focuses on evaluating whether the results are robust to MAR assumptions implicit in MI and ML analytic approaches [9].

Other methods have been proposed to assess the influence of MNAR data. The combination of high attrition and data that are potentially MNAR is not unique to mHealth research, and several approaches have come from the addiction field [19,20]. A classic example of MNAR data occurs in smoking cessation research, where individuals who drop out of the study are more likely to have returned to smoking. Historically, a widely used approach to handling missing smoking cessation data is simply to assume that missingness equals smoking [14]. This approach is considered conservative and is arguably preferred over treating the missing values as MAR or MCAR. However, assuming missing equals smoking can also introduce bias [21]; for example, if missingness is strongly related to group assignment in the context of an RCT and not all participants who drop out, in fact, return to smoking (eg, higher attrition in the waitlist vs nicotine patch condition).

Hedeker et al [14] offered a sensitivity analysis approach for evaluating the impact of MNAR on study results within the context of smoking cessation that could be adapted for use in mHealth research. Specifically, Hedeker et al [14] recommend evaluating the sensitivity of results to varying assumptions about the smoking status for those with missing data [14]. Models range, for example, from assuming a perfect association between missingness and smoking (ie, missing=smoking) to assuming that the odds of smoking for an individual with missing data are between 2 and 5 times higher than those with nonmissing data [14,22]. If the results are robust to variations in the assumed value of missing data, one can be more confident that the potential MNAR does not undermine the findings. If the results change, one can also characterize the point at which this occurs (eg, shifting from statistical significance to nonsignificance). Similar approaches have been proposed in other fields as well (eg, cost-effectiveness analyses) and incorporated into a broader MAR framework (eg, MI [23]).

Despite a longstanding acknowledgment that missing data are common in mHealth research [1], to our knowledge, there has not been a systematic investigation of the nature of missing data (ie, MAR vs MNAR) and no recent evaluation of the ways in which study authors are handling missing data (for an older review of missing data analysis techniques in internet-based interventions for anxiety and depression, refer to Christensen et al [2]). As noted, it is unfortunately not possible to definitively determine whether missing data are MAR or MNAR [8]; by definition, one cannot establish an association between the likelihood of a missing value and the unobserved value itself. Some readers may be familiar with Little [24] MCAR test, which is designed to evaluate the likelihood of MCAR across a data set. Although it is tempting to consider this as a reliable option for establishing missingness as MCAR, it has a number of substantial drawbacks, including low power (which can lead to failure to reject the null hypothesis that data are MCAR), unlikely and untestable assumptions (eg, shared covariance matrix among miss data patterns), and failing to identify specific variables that violate MCAR (ie, providing only an omnibus test [8]). In the absence of a method for determining whether data are MNAR, one could argue that it is incumbent upon researchers to consider whether their handling of potentially MNAR data yields biased results.

As noted above, a pattern of missingness that may be suggestive of MNAR in mHealth research is when missingness is higher in an active condition relative to a passive (eg, waitlist) control group. The context of an RCT is important for making this claim. Random assignment should produce groups balanced on all relevant covariates at baseline, including those that would predict drop out [25]. As attrition would be caused, at least in part, by group assignment (ie, active vs waitlist), it is, therefore, important to speculate on the primary mechanism by which treatment creates missingness. In the context of mHealth, one could easily imagine that active participants are more likely to drop out because of the increased burden associated with their intervention. Presumably, participants who find the burden of remaining engaged to exceed the benefits (or lack of benefits) they are experiencing may be most likely to drop out. Likewise, participants who experienced adverse reactions to the intervention itself would be more likely to drop out. In both instances—participants failing to realize benefits or experiencing adverse reactions—missing posttreatment data are likely MNAR, with unobserved scores on average reflecting less improvement than observed scores. Regardless of the specific cause, the meaning of missingness in the active condition will almost certainly not be equivalent to the missingness in waitlist control. This makes it problematic to treat missing data as reflecting the same outcomes as others in their respective groups, which is precisely what MAR methods do.

A recent meta-analysis of attrition in RCTs testing smartphone-based mental health interventions [3] found evidence consistent with this potential source of MNAR data. Linardon and Fuller-Tyszkiewicz [3] noted that active participants were significantly more likely to drop out of the RCTs than the passive control group participants (odds ratio [OR] 1.87, 95% CI 1.45-2.41, across all follow-up time points). In contrast, this differential attrition was not observed when an active control condition was used (OR 1.13, 95% CI 0.91-1.42). As Linardon and Fuller-Tyszkiewicz's [3] study was not primarily focused on differential attrition, they did not further explore the possibility of MNAR or its implications, nor did they conduct standard meta-analytic sensitivity analyses for this specific effect (eg, trim-and-fill adjustment [26]). It would be valuable to extend this finding by systematically evaluating how differential attrition is handled statistically in these mHealth RCTs and examining study design features associated with higher rates of differential attrition (ie, meta-analytic moderators).

In addition to further clarifying how differential attrition and potential MNAR are handled within the mHealth literature, there is a need to understand the potential implications of MNAR for study outcomes. Selection models and pattern-mixture models are two promising approaches. Sensitivity analyses such as those recommended by Hedeker et al [14] for smoking cessation could also be readily adapted for mHealth research.

This Study

This study has 2 primary aims. The first is to systematically review the analytic methods used to address missingness in a portion of the mHealth literature that has previously shown indications of potential MNAR. We examined 36 RCTs drawn from Linardon and Fuller-Tyszkiewicz's [3] recent meta-analysis of smartphone-based mental health interventions that compared active interventions and passive controls. To examine statistical moderators of differential attrition, we coded attrition and study design features. We then cataloged how missing data were handled within these trials, focusing on whether the statistical approaches could handle MAR or MNAR data.

Our second aim is to present methods for evaluating the effects of MNAR that may be relevant to mHealth research. We illustrate the value of sensitivity analyses by applying an MI-based pattern-mixture model along the lines of Hedeker et al [14], as well as a simpler fixed-value replacement sensitivity analysis as examples of informative methods for evaluating the impact of MNAR. To illustrate these approaches, we use data drawn from a recent RCT of a smartphone-based mental health intervention comparing 2 active conditions with a waitlist control group [27].

Methods

Assessment of MNAR and Systematic Review of Missing Data Analytic Approaches

To evaluate study design features associated with differential attrition and the methods used to handle missing data, we reanalyzed and systematically reviewed RCTs included in the meta-analysis of attrition in smartphone-based mental health interventions by Linardon and Fuller-Tyszkiewicz [3]. This meta-analysis is recent and includes a reasonably large sample of RCTs (n=36 studies) that compared active treatment with a passive control condition (ie, waitlist or no treatment). We coded the completer and drop out sample sizes for both active and passive conditions at posttreatment to characterize differential attrition. These values were then converted to ORs using standard meta-analytic methods [26]. Log ORs and their variance were then aggregated using a random effects meta-analysis, weighted as typical using inverse variance [26] in the metafor R package (R Core Team). As ORs and the variance of ORs cannot be computed for cells with zeros, we conducted analyses using the Peto method [28], as recommended in the Cochrane handbook [29]. We also conducted analyses by adding a continuity correction for instances of empty cells (ie, 0.5 added to all cells in a study with an empty cell [30]). Heterogeneity of effect sizes was characterized using I^2 (ie, proportion of effect size variance that occurs between studies) and interpreted based on Higgins et al [31]. We assessed the potential influence of outliers by conducting a *leave-one-out* analysis in the metafor package [32] and using the *find.outliers* function in R [33] that excludes effect sizes whose CI do not overlap with the omnibus effect size CI.

We systematically reviewed several features of the included studies. These included the overall sample size, overall dropout rate, whether potential differential attrition was statistically evaluated (ie, comparing dropout rates for active vs passive conditions), whether differential attrition was detected, the approach used for handling missing data, whether a modern MAR data analytic approach was used (ie, MI or ML), and whether a sensitivity analysis was conducted to evaluate the potential impact of MNAR data. To evaluate whether these study characteristics were linked with differential attrition, we tested them as moderators [26]. All analyses were conducted using R [34].

MNAR Sensitivity Analysis

We used data from a recently conducted RCT testing a smartphone-based mental health intervention [27] to illustrate 2 sensitivity analysis approaches for MNAR data. As many mHealth RCTs include pre- and posttreatment assessments on a continuous variable, we apply these sensitivity analyses using data of this kind. In this study, 2 versions of an active smartphone-based meditation intervention were compared with a waitlist control on changes in psychological distress for 8 weeks (n=343). The original RCT included 3 time points (pretest, midtreatment, and posttreatment), and the primary models used multilevel modeling with ML estimation. However, in keeping with the possibility of MNAR data, attrition was higher in the active intervention than the waitlist (OR 2.10, 95% CI 1.34-3.33).

The first approach is a variant of the pattern-mixture model [23]. First, one conducts MI, imputing missing values based on available data (eg, pretest scores and demographics). Code in Multimedia Appendix 1 implements this in R using the jomo [35], mitools [36], and mice [37] packages with 100 multiply imputed data sets. It is worth noting that a limitation of MI in this context is the likely simulation of a positive treatment effect in the missing outcomes (assuming a positive treatment effect is seen in the observed outcomes), which may not be correct in the presence of MNAR. Thus, we next modify the imputed (ie, previously missing) posttest values using an *offset* parameter representing varying MNAR conditions. In our example, we assume progressively worse outcomes for those with missing posttest values. As a lower distress score is better, we add positive constants defined in relation to the residual SD from a model predicting posttest scores controlling for pretest scores and group status. As the added positive constant increases, the assumed outcome for missing observations becomes progressively worse. To aid in interpretability, we followed Cohen [38] effect size convention and added this value multiplied by 0.20, 0.50, 0.80, 1.10, and 1.40 to the multiply imputed values for cases with missingness. For example, the deviation applied for the 0.20 condition is as follows:

$$\text{Missing} = \text{Multiply imputed value} + 0.20 \times \text{SD}_{\text{Model}} \quad \textbf{(1)}$$

A possible limitation of this approach is that, in the possible absence of useful covariates in predicting missing outcomes, all missing observations will be generated with large SDs, implying a high degree of uncertainty in the missing outcomes. Thus, even when introducing the offset parameter following the pattern-mixture strategy, the observed variability in the missing observations will still be large. To the extent that we should not confuse lack of knowledge about missing outcomes with actual variability in the missing outcomes, a sensitivity analysis that also considers a fixed-value replacement for missing observations can be useful. Therefore, we also applied a second sensitivity analysis approach outside the context of MI. Consistent with our strategy, this second approach focuses on estimating residualized change scores, although simple change scores could also be used. Once residualized change scores are imputed for missing cases, nonparametric tests (eg, Wilcoxon signed-rank test) can then be conducted using these values to compare changes in the active and passive conditions while avoiding statistical drawbacks associated with conducting parametric tests using single imputed data (eg, artificially deflating SE by treating imputed values as if they were observed values [8]). Similar to the approach described above, to evaluate the influence of potential MNAR data, we tested varying assumptions about the meaning of missing posttest data from complete case analysis to a worst-case scenario. The first analysis assumes that the missing data are MCAR and uses complete cases.

$$\text{Complete case analysis: Missing=NA} \quad \textbf{(2)}$$

For the worst-case scenario, the missing data were assumed to reflect the worst possible observed outcome. Residualized change is operationalized as the observed posttreatment score minus the predicted posttreatment score based on pretreatment. For an outcome such as distress, in which lower values are preferred (ie, lower distress), a larger (ie, more positive) residual indicates a smaller decline in distress (for negative values), or even an increase in distress over time (for positive values). For an outcome in which higher scores were better (eg, well-being), one would simply reverse this approach (ie, replace missing values with the minimum observed residual). In our example, the worst-case scenario replaces the missing values with the maximum value of the observed residualized change scores:

$$\text{Worst-case scenario: Missing=Maximum}_{\text{Residual}} \quad \textbf{(3)}$$

We then evaluated possibilities between the complete and worst-case scenarios, with missing values imputed to be 0.20, 0.50, and 0.80 SD from the mean residualized change score. These specific values were chosen to reflect small, medium, and large deviations based on Cohen [38] guidelines. Again, as a lower score is better for distress, these deviations were added to the mean residualized change score (the mean residual is expected to be zero but is included here for the sake of completeness):

$$\text{Small MNAR deviation: Missing} = \text{Mean}_{\text{Residual}} + 0.20 \times \text{SD}_{\text{Residual}} \quad \textbf{(4)}$$

$$\text{Medium MNAR deviation: Missing} = \text{Mean}_{\text{Residual}} + 0.50 \times \text{SD}_{\text{Residual}} \quad \textbf{(5)}$$

$$\text{Large MNAR deviation: Missing} = \text{Mean}_{\text{Residual}} + 0.80 \times \text{SD}_{\text{Residual}} \quad \textbf{(6)}$$

For example, psychological distress in the RCT by Goldberg et al [27] was a composite of 3 measures assessing depression, anxiety, and stress, which were combined into a single measure and scaled to z units (ie, mean 0, SD 1). The mean residualized change in psychological distress was 0 (SD 0.65), and the maximum residualized change in psychological distress was 2.3. Therefore, the worst-case scenario replaced all the missing residualized change scores of 2.3. In the midrange scenarios, missingness was replaced with a small deviation from the mean (0 + 0.2 × 0.65 = 0.13), a medium-sized deviation from the mean (0 + 0.50 × 0.65 = 0.33), and a large deviation from the mean (0 + 0.80 × 0.65 = 0.52). Wilcoxon signed-rank tests compared the rank sum for the active and passive conditions based on the complete case analysis and the 4 scenarios. All analyses were conducted using R [34]. Deidentified data [39] and the R code necessary for conducting the sensitivity analyses are included in Multimedia Appendix 1.

Results

Assessment of MNAR and Systematic Review of Missing Data Analytic Approaches

Linardon and Fuller-Tyszkiewicz’s [3] review included 36 RCTs that compared one or more active conditions with a waitlist control condition. Intention-to-treat and completer sample sizes, along with study characteristics related to missing data analysis, are included in Table 1. The average sample size per study, combined across active and passive conditions, was 143.53 (SD 118.66). Average attrition rates were numerically higher in the active condition (23.32%, SD 19.88%) than in the passive condition (15.36%, SD 15.51%), and 2 studies reported no attrition [40,41]. Among the 34 studies with attrition, a minority (11/34, 32%) statistically compared attrition rates between active and passive conditions. A total of 6 studies detected differential attrition, in all cases reporting higher attrition in the active conditions relative to the passive conditions.

Table 1. Attrition rates and study design characteristics.

Study	Tx[a] ITT[b]	Tx drop	WL[c] ITT	WL drop	Diff[d]	Method[e]	Multiple imputation	Maximum likelihood
Bakker et al [42]	234	146	78	25	N/A[f]	ANOVA[g]	Yes	No
Bidargaddi et al [5]	192	106	195	88	Yes, higher in active	*t* tests	Yes	No
Bostock et al [43]	128	5	110	4	N/A	ANOVA	No	No
Carissoli et al [40]	20	0	18	0	N/A	ANOVA	N/A	N/A
Champion et al [44]	38	9	36	3	No	MLM[h]	Yes	Yes
Enock et al [45]	206	38	36	0	N/A	MLM	No	Yes
Faurholt-Jepsen et al [46]	39	6	39	5	N/A	MLM	No	Unclear[i]
Hall et al [47]	76	34	25	13	N/A	MLM	No	Unclear
Horsch et al [48]	74	29	77	15	N/A	MLM	Yes	Unclear
Ivanova et al [49]	101	20	51	4	N/A	MLM	No	Yes
Kahn et al [50]	80	1	80	0	N/A	*t* tests	No	No
Krafft et al [51]	67	15	31	5	N/A	MLM	No	Yes
Kristjansdottir et al [52]	70	23	70	33	N/A	*t* tests	No	No
Kuhn et al [53]	62	11	58	6	No	ANOVA	Yes	No
Lee and Jung [54]	102	25	104	18	N/A	ANOVA	No	No
Levin et al [55]	12	0	11	0	N/A	MLM	No	Unclear
Levin et al [56]	59	13	28	5	No	MLM	No	Unclear
Lüdtke et al [11]	45	10	45	6	No	ANOVA	Yes	No
Lukas and Berking [57]	16	2	15	2	N/A	ANOVA	No	No
Ly et al [58]	36	3	37	2	N/A	MLM	No	Yes
Ly et al [41]	14	0	14	0	N/A	MLM	No	Yes
Marx [59]	46	2	50	0	N/A	ANOVA	No	No
Miner et al [60]	25	2	24	3	N/A	ANOVA	Yes	No
Moëll et al [61]	29	3	28	1	N/A	ANOVA	No	No
Oh et al [62]	39	1	20	4	N/A	ANOVA	No	No
Pham et al [63]	31	14	32	7	N/A	ANOVA	No	No
Proudfoot et al [64]	242	116	230	32	Yes, higher in active	MLM	Yes	Yes
Roepke et al [65]	190	152	93	57	Yes, higher in active	MLM	No	Yes
Rosen et al [66]	57	17	55	7	Yes, higher in active	MLM	No	Yes
Schlosser et al [67]	22	3	21	0	N/A	ANOVA	No	No
Stjernsward and Hansson [68]	196	60	202	42	N/A	ANOVA	Yes	No
Stolz et al [69]	60	18	30	7	No	MLM	Yes	yes
Tighe et al [70]	31	2	30	0	N/A	ANOVA	No	No
van Emmerik et al [71]	191	111	186	45	Yes, higher in active	MLM	Yes	Unclear
Versluis et al [72]	46	9	42	3	Yes, higher in active	MLM	No	Unclear
Yang et al [73]	45	3	43	4	N/A	ANOVA	No	No

[a] Tx: active treatment conditions.

[b] ITT: intention-to-treat sample size; drop=attrition at posttreatment assessment.

[c] WL: waitlist (or no treatment control condition).

[d] Whether differential attrition was tested and, if so, whether a between-group difference was detected.

[e] Primary data analysis method.

[f] N/A: not applicable (because of lack of missing data or differential attrition test not conducted).

[g] ANOVA: analysis of variance or related method (eg, analysis of covariance).

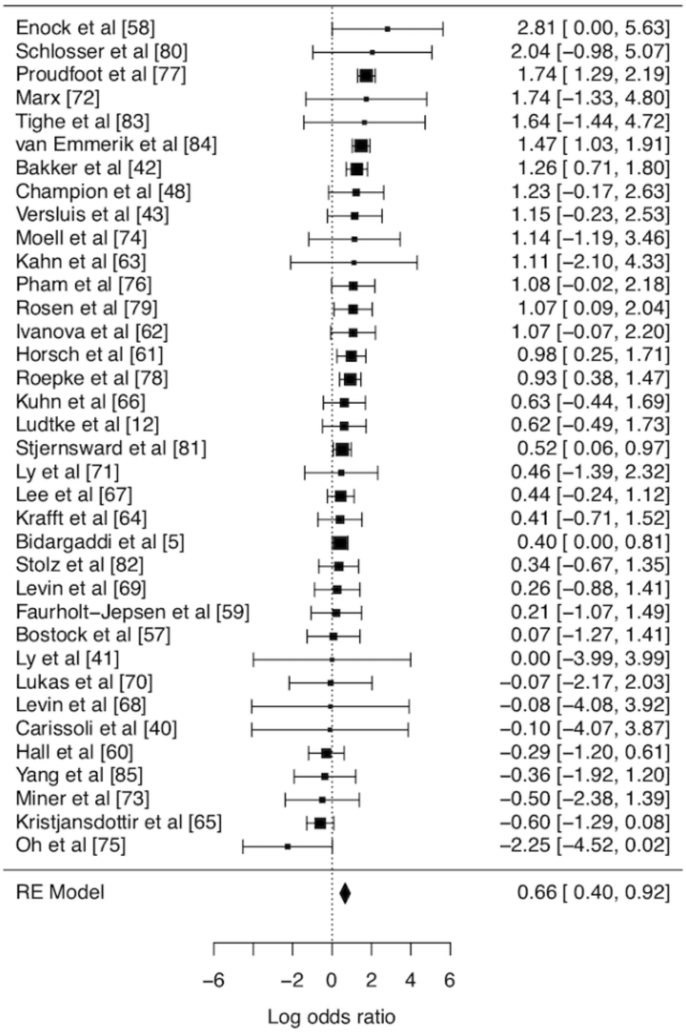
[h] MLM: multilevel model.

[i] Unclear whether multiple imputation estimator was used.

Most studies used multilevel modeling (17/36, 47%) or a variant of analysis of variance (eg, analysis of covariance, multivariate analysis of variance; 16/36, 44%) as the primary analytic approach, with 8% (3/36) of studies using a *t* test. Half (18/36, 50%) of the studies used ML or MI to handle missing data. MI was used in 31% (11/36) of studies, with multiple imputed data sets then analyzed using analysis of variance or *t* tests [5,42]. ML was used in combination with multilevel modeling in 28% (10/36) of the studies. An additional 19% (7/36) of studies used multilevel modeling but did not specify the estimator [72]. No study conducted a sensitivity analysis to evaluate the potential impact of MNAR data.

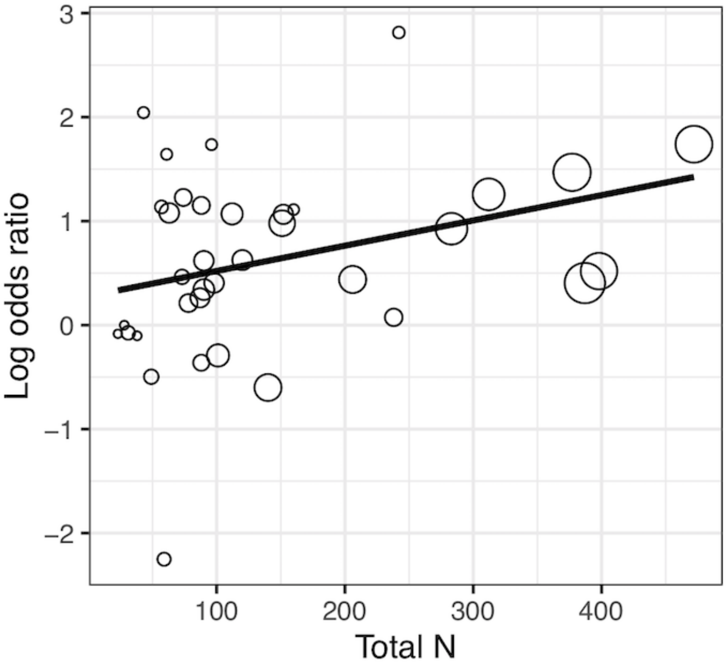
Consistent with Linardon and Fuller-Tyszkiewicz [3], the results of our reanalysis provided a clear indication of differential attrition, with participants randomized to the active conditions approximately twice as likely to drop out relative to those in passive conditions (OR 1.94, 95% CI 1.50-2.51, using the Peto method; OR 2.22, 95% CI 1.93-2.54 using a continuity correction; both *P* values are <.001; Figure 1). The heterogeneity was moderate (I^2=53.85%, 95% CI 19.46-71.09). The results were robust to *leave-one-out* analyses (OR range 1.82-2.10; all values of *P*<.001). The *find.outlier* function detected 4 outliers. Results were similar with these studies removed (OR 1.91, 95% CI 1.58-2.32; *P*<.001).

Figure 1. Forest plot displaying results of the meta-analysis. Effects sizes are in log-odds units, with larger values indicating higher attrition in active conditions relative to passive conditions. The size of points indicates relative weight in the meta-analysis (ie, inverse variance). RE: random effects.



Several potential moderators were assessed using a meta-regression analysis. Active conditions were more likely to show higher attrition than passive conditions as the overall sample size increased (B=0.0022, 95% CI 0.0005-0.0039; note that all meta-regression coefficients are in log OR units; *P*=.01; Figure 2). Higher overall attrition was not associated with differential attrition (B=0.57, 95% CI −0.89 to 2.03; *P*=.45). Studies with higher differential attrition were marginally more likely to test for differences in attrition rates between active and passive conditions (B=0.49, 95% CI −0.002 to 0.99; *P*=.05, where testing=1 and not testing=0). There was no association between differential attrition rate and the likelihood of detecting differential attrition (B=0.55, 95% CI −0.19 to 1.29; *P*=.15, where detecting a difference=1 and not detecting a difference=0). The use of a modern missing data analysis method (ie, ML or MI) was associated with higher rates of differential attrition (B=0.73, 95% CI 0.25-1.2; *P*=.003, where use of ML or MI=1, no use of ML or MI=0).

Figure 2. Results of meta-regression indicating that larger studies are associated with higher rates of differential attrition (ie, higher attrition in active vs passive conditions). Points are displayed relative to their weight in the meta-regression model (ie, inverse variance).



MNAR Sensitivity Analysis

Selection and pattern-mixture models are 2 valuable modeling strategies for handling MNAR (see Multimedia Appendix 2 [8,74,75] for a brief discussion of these methods and their limitations). New strategies and extensions of these approaches are continually being developed (eg, the index of local sensitivity to nonignorability [76]). However, many mHealth researchers may not be familiar with these methods. Selection models, in particular, require the missing data mechanism to be specified, which can be difficult to do. Moreover, a pattern-mixture approach of the kind described above often reflects a larger degree of uncertainty in the missing observations, an uncertainty that should not be confused with the presence of known variability in the missing outcome. Therefore, rather than abandon attempts to assess the potential impact of MNAR, mHealth researchers could consider approaches that simply make specific assumptions regarding anticipated outcomes for missing observations (ie, fixed-value replacement). By examining the estimated treatment effects in the presence of specific assumed outcomes for missing observations, we can similarly provide some insight into the degree to which varying missingness assumptions impact study results [14]. We illustrate both the pattern-mixture model and fixed-value replacement approaches using data drawn from the RCT by Goldberg et al [27].

Of the 343 participants randomized, 228 (66.5%) were assigned to 1 of the 2 active conditions, and 115 (33.5%) were assigned to the waitlist control. Consistent with the possibility of MNAR, noncompletion of posttreatment assessments was higher in the active condition (137/228, 60.1%) than in the waitlist condition (48/115, 41.7%; OR 2.10, 95% CI 1.34-3.33; *P*=.001). Goldberg et al [27] primary analyses used all 3 time points in multilevel models with ML estimation. The results indicated a steeper decline in psychological distress for the active conditions relative to the waitlist (time × group interaction; *P*<.001). Here, we examine how this result changes based on varying MNAR scenarios using either an MI-based pattern-mixture model approach [23] or a fixed-value replacement sensitivity analysis approach.

Table 2 shows the estimates of the effect of group status on posttest distress, controlling for pretest distress across varying MNAR conditions within the pattern-mixture model framework. As the positive constant added (ie, offset parameter) increases, increasingly worse outcomes are assumed for the missing observations. Those in the active group continued to show larger declines in distress until imputed posttest distress scores were offset by a value of 1.10 or greater residual SD. Figure 3 depicts the impact of these varying MNAR conditions. The first panel displays the MAR-based estimates provided by MI, with imputed values closely following the trajectory of the respective groups. As MNAR conditions vary, the trajectories for imputed values become increasingly divergent from the observed scores, including the point that they reflect worsening scores with time.

Table 2. Results of pattern-mixture model sensitivity analysis based on multiple imputation[a].

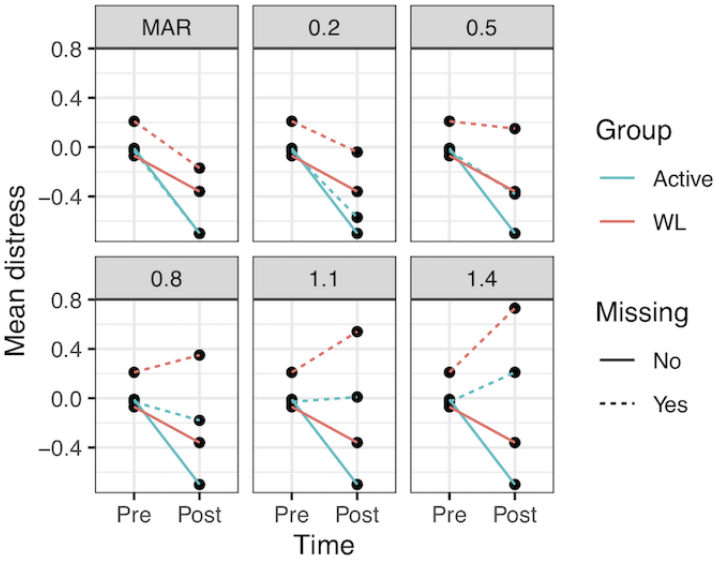
Model	Estimate[b]	*P* value
MAR[c]	−0.34	.002
0.20	−0.31	.004
0.50	−0.28	.01
0.80	−0.24	.03
1.10	−0.20	.08
1.40	−0.17	.16

[a] Models are based on varying assumptions regarding the meaning of missingness. Multiply imputed posttest values based on 100 imputations are offset [23] by varying amounts (ie, 0.20, 0.50, and residual SD).

[b] Coefficient for active group status (vs waitlist) predicting posttest distress scores controlling for pretest distress scores pooled across imputed data sets.

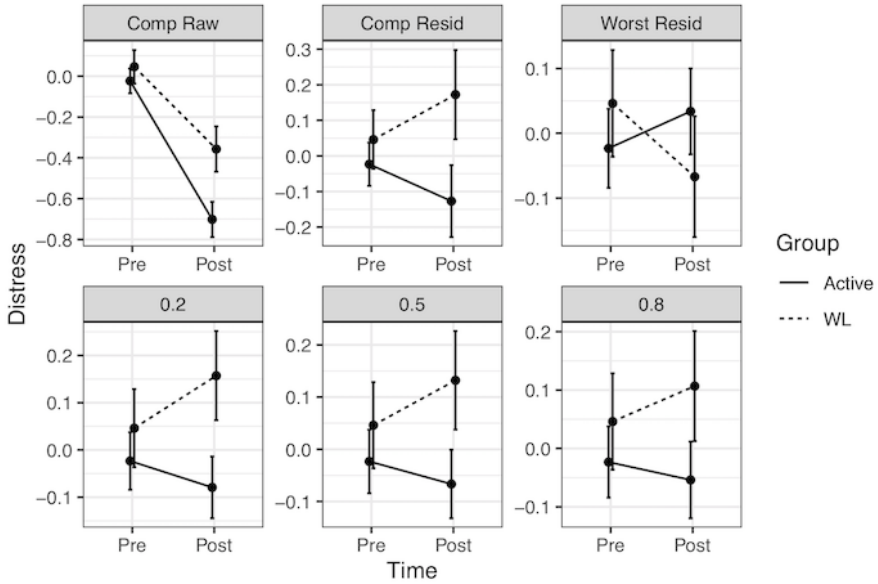
[c] MAR: missing at random (with no offset applied to posttest values).

Figure 3. Pre- and posttreatment scores for active and passive conditions with varying constant offset parameters added to multiply imputed values for missing outcomes under conditions of missing not at random (ie, Missing). Values are in z-score units, scaled by distress at baseline (mean 0, SD 1). Panels illustrate trajectories with offsets ranging from 0.2 to 1.4 residual SD. The missing at random panel represents values derived using multiple imputation with no offset applied. MAR: missing at random; WL: waitlist.



We now turn to the results of the fixed-value replacement sensitivity analysis. Figure 4 visually depicts the impact of MNAR conditions on the trajectories of pre-post change for the active and passive groups using this approach. The first 2 panels (Comp Raw and Comp Resid) display changes for completers only (in raw units and residualized change units, respectively). However, if MAR is violated in the way hypothesized above, one would expect the trajectory for unobserved active group participants to be worse than the observed active group scores (ie, following a trajectory more similar to the passive condition). If the missing data are consistent with MAR, this adjusted trajectory can be adequately recaptured with observed data (eg, baseline variables), allowing unbiased estimation using ML and MI. However, in the case of MNAR, the likelihood of missingness depends upon the unobserved value itself, making it impossible to recapture from available data alone. The subsequent panels (Worst Resid, 0.20, 0.50, and 0.80) display the impact of varying assumptions about the meaning of missing values. As can be seen in the Worst Resid panel, assuming the worst observed outcome for those with missing data reverses the direction of effect, with control group participants now showing more improvement than active participants. One can see how the gap in outcomes between active and passive condition participants narrows as increasingly strong assumptions are made regarding the degree to which missing values deviate from observed values. As missingness was more prevalent in the active conditions, these modifications exerted a stronger influence on the change in the active group.

Figure 4. Pre- and posttreatment scores for active and passive conditions under varying missing not at random conditions using fixed-value replacement of missings. Pretreatment values represent z-scaled distress at baseline (mean 0, SD 1). Posttreatment values vary across plots. For Comp Raw, posttreatment values are posttreatment distress scaled based on baseline distress. Subsequent plots display residualized change scores z-transformed at posttreatment to aid in visual interpretation of relative, between-group pre-post change. Comp Resid computed posttreatment as baseline plus residualized change scores for completers only. Worst Resid replaced missing posttreatment Comp Resid values with the lowest improvement in distress. Subsequent figures (0.2, 0.5, 0.8) replaced missing posttreatment Comp Resid values with values 0.2, 0.5, and 0.8 SD worse than the mean residual. Comp: completer; Resid: residualized change; WL: waitlist.



For null hypothesis testing purposes, we used nonparametric tests of mean residualized change scores. Consistent with the multilevel modeling results [27], the Wilcoxon signed-rank test favored the active conditions in the completer sample (mean ranks 69.11, 93.61, SD 43.01 and 45.90, for active and passive conditions, respectively, *P*<.001, where a lower rank indicates a larger decline in distress; Table 3). In the worst-case scenario, the direction of the mean rank difference flipped, now favoring the passive condition, although only marginally significantly (*P*=.08). Mirroring Figure 4, the influence of the varying missingness assumptions is apparent in Figure 5. The gap between active and passive conditions narrows, as missing data are assumed to reflect poorer and poorer outcomes. The pattern specifically indicates that statistical significance persists when missing values are assumed to be 0.20 above the mean residual but not 0.50 or higher. This result differs slightly from Goldberg et al [27], who detected statistical significance at an offset of 0.50. The discrepancy is because of Goldberg et al [27] calculating the SD for the residual without the group variable in the model. We recommend the inclusion of the group variable in the model, as the resultant SD is presumably more conservative, based on the assumption that an intervention increases the SD. Therefore, these results can be interpreted as robust to MNAR missing, in which the unobserved values deviate from the observed values only to a small degree, but not when showing moderate or larger deviations.

Table 3. Results of fixed-value replacement sensitivity analysis.

Group and model	Sample size, n (%)	Mean rank (SD)	SE	*P* value[a]
Active				
Comp[b]	91 (39.9)	69.11 (43.01)	4.51	<.001
Worst[c]	228 (100)	178.1 (93.05)	6.16	.08
0.20[d]	228 (100)	161.89 (83.25)	5.51	.004
0.50[d]	228 (100)	165.19 (84.25)	5.58	.05
0.80[d]	228 (100)	168.52 (86.01)	5.70	.32
Waitlist				
Comp	67 (58.3)	93.61 (45.9)	5.61	N/A[e]
Worst	115 (100)	159.9 (85.62)	7.98	N/A
0.20[d]	115 (100)	192.05 (102.27)	9.54	N/A
0.50[d]	115 (100)	185.5 (102.26)	9.54	N/A
0.80[d]	115 (100)	178.9 (100.34)	9.36	N/A

[a] *P* value from Wilcoxon signed-rank test comparing active and passive conditions across varying missingness assumptions.

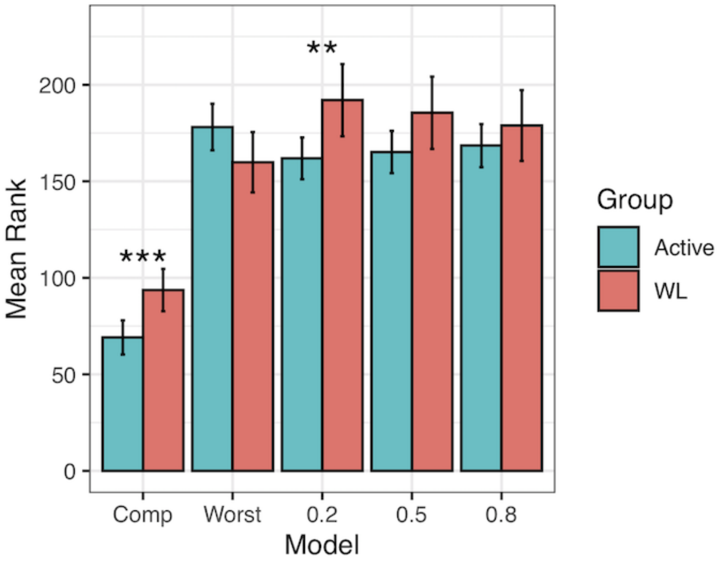
[b] Comp: completer sample.

[c] Worst: worst-case scenario, which assumed missing values are equivalent to the worst outcome (ie, smallest change in distress).

[d] 0.20, 0.50, 0.80: missing values assumed to be 0.20, 0.50, or 0.80 SDs worse than the mean residualized change score.

[e] N/A: not applicable.

Figure 5. Results of Wilcoxon signed-rank test using a fixed-value replacement sensitivity analysis across varying missing not at random conditions. A lower mean rank indicates larger relative decreases in distress. Comp: completer sample; Worst: worst-case scenario which assumed missing values are equivalent to the worst outcome (ie, the smallest change in distress); 0.2, 0.5, 0.8: missing values assumed to be 0.2, 0.5, or 0.8 SD worse than the mean residualized change score; error bars: 1.96×SE; WL: waitlist. **P*<.05, ***P*<.01, ****P*<.001.



Discussion

Principal Findings

This study had two primary aims: to systematically evaluate the handling of a potential source of MNAR data in mHealth research—differential attrition—and advocate for sensitivity analyses as a family of strategies that might be used to assess the impact of MNAR data. At the broadest level, results suggest that MNAR data are likely to be a problem in mHealth research and one that, to date, has not been adequately addressed. As reported by Linardon and Fuller-Tyszkiewicz [3], the substantially higher attrition in active relative to passive conditions in RCTs testing smartphone-based mental health interventions is marked; active participants were approximately twice as likely to drop out of the study. Although it is impossible to say what the missing posttreatment data would have shown had it been collected, it is plausible that those dropping out from the active conditions were less likely to have benefited from the mHealth intervention (or at least that the benefits they were experiencing did not outweigh the costs of remaining in the study). Thus, the observed values may overestimate treatment effects for those under active conditions. Given that the likelihood of missingness is related to the unobserved values themselves, these data would be MNAR.

Although patterns of attrition consistent with potential MNAR data were detected in the literature as a whole, only a minority of the included studies tested for differential attrition. However, in keeping with the literature-wide pattern of differential attrition, 6 of the 11 studies comparing attrition rates between active and passive conditions detected higher attrition in the active conditions, whereas the remaining 5 studies failed to detect a difference. Despite the possibility of MNAR data, none of the 36 studies directly assessed the potential influence of MNAR data on the study results. Half of the studies employed other modern missing data methods, such as MI or ML. These approaches have many strengths and are certainly preferred over historical approaches for handling missingness (eg, last observation carried forward and complete case analysis [9]). Encouragingly, it appears that less sophisticated missing data analysis techniques (eg, last observation carried forward [2]) are being replaced by modern methods. However, both MI and ML rely on the assumption that data are MAR; therefore, missing values can be reliably determined based on measured variables. Importantly, they are not robust to MNAR [8].

Moderator analyses further characterized the correlates of differential attrition. The results indicated that differential attrition was more likely to occur in larger studies. Unfortunately, this association could produce a pernicious source of bias within the literature, as larger studies are presumably the ones most looked to when evaluating evidence of efficacy and are likely to carry more weight in meta-analyses examining efficacy. Interestingly, studies with higher differential attrition were only marginally more likely to assess differential attrition. It may be that differential attrition is simply not recognized or acknowledged as a potential concern worth assessing, even when dropout rates differ. Somewhat counterintuitively, studies with higher differential attrition were not more likely to detect differential attrition when assessed. This lack of association could be because of the limited statistical power for the moderator test itself [77], as only 11 studies tested for differential attrition. Statistical power may also be low in primary studies themselves. For example, Champion et al [44] did not detect differential attrition in their sample of 74 participants, although active participants were 3.42 times more likely to drop out of the active condition relative to the passive condition. It appears that researchers are more likely to use modern missing data analysis techniques (ML/MI) when differential attrition is higher, which is preferred to techniques that are not robust to even MAR data (eg, complete case analysis and last observation carried forward). Nonetheless, these techniques are not capable of eliminating the bias associated with MNAR data.

Perhaps the most notable finding of our review is that none of the included studies conducted a sensitivity analysis to evaluate the potential influence of MNAR data on study findings. Although several meta-analyses suggest that smartphone-based mental health interventions produce benefits relative to waitlist control conditions [78-80], the lack of sensitivity analyses coupled with literature-wide differential attrition makes the apparent efficacy more tenuous.

The primary aim of this study is to encourage mHealth researchers to consider sensitivity analyses to assess the potential impact of MNAR missingness, particularly when differential attrition is present. Several modern techniques exist for evaluating the potential impact of MNAR missingness, including selection models and pattern-mixture models that have been discussed. As most of these methods have limitations (eg, they are heavily influenced by untestable assumptions) and may not be within the current analytic repertoire of many mHealth clinical trialists, we presented an MI-based pattern-mixture model sensitivity analysis approach adapted from smoking cessation research [14,22] as well as a fixed-value replacement sensitivity analysis approach as examples of more user-friendly strategies for evaluating the impact of MNAR data. These methods are fairly straightforward to implement using a continuous outcome variable assessed at pre- and posttreatment—a typical situation for mHealth research [80,81]—and move beyond the traditional MAR methods currently emphasized in mHealth research. An attractive feature of these sensitivity analyses is that one can visually and statistically evaluate the impact of varying missingness assumptions on the pattern of findings. As these assumptions would only apply in cases of missing data, they would have a minimal impact on the results when missingness is low (eg, <5% [9]). As expected, the actual impact of varying MNAR assumptions will be sensitive to other patterns in the data (eg, trajectories of change for waitlist control participants because of regression to the mean or natural history). Thus, they do not imply a particular direction of influence but rather evaluate a range of possible impacts based on deviations from the observed data.

It is worth noting that the 2 sensitivity analysis approaches illustrated in this study provided somewhat discrepant conclusions regarding the degree to which data from Goldberg et al [27] were robust to MNAR conditions. This fact highlights the value of sensitivity analyses and the importance of authors using various approaches and assumptions to evaluate the strength of their findings. These differences are also illuminating. In particular, the MI-based pattern-mixture model approach suggested that the results were robust to MNAR deviations that were large (ie, 0.80) but not larger, whereas the simpler sensitivity analysis approach indicated that the results were not robust above small deviations (ie, 0.20). Figure 3 illustrates a plausible explanation for this: the MI-based approach makes the initial assumption that missing values are similar to observed values unique to each group. Thus, the fact that the active group improved overall produced improvement in the imputed change for missing active participants. In contrast, the fixed-value replacement approach did not adjust the expected residualized change scores based on the group status. We contend that both approaches may provide a valuable perspective on MNAR sensitivity and should simply be interpreted in light of their underlying assumptions.

Limitations and Future Directions

This study had several important limitations. The first and broadest limitation is that we cannot definitively conclude that the observed differential attrition necessarily results in MNAR data. It is possible that remaining in the study was because of factors unrelated to changes in study endpoints (ie, distress). Likewise, drop outs in the active group could have been because of participants not using the smartphone app and losing interest in the study because their psychological symptoms had already improved (as can be the case in psychotherapy [82]), which could produce an MNAR bias in the opposite direction (ie, missing values are better, not worse). As is typical for research on missing data, the data necessary to test for MNAR are by definition missing. The methods proposed here could certainly be extended to evaluate potential *best-case scenarios*, in which missing observations reflect better rather than worse outcomes or when missingness has different meanings depending on group assignment (eg, worse outcomes for active conditions but better outcomes for passive condition). Second, we only evaluated the degree and correlates of differential attrition in a small subset of the large and rapidly growing mHealth literature. It is possible that researchers are improving their ability to retain study participants and adherence strategies being investigated [6,83] may be decreasing attrition in the active conditions. Future reviews may see less evidence of this potential source of MNAR data. Similarly, there are mHealth RCTs that conducted sensitivity analyses to evaluate MNAR (eg, pattern-mixture models [84]), even though none of the 36 RCTs with passive controls we evaluated did so. Third, we focused only on differential attrition in smartphone-based RCTs. It is conceivable that higher attrition in active than passive conditions is somehow idiosyncratic to this delivery platform. An important future direction would be to evaluate differential attrition in other mHealth delivery formats (eg, internet-based interventions). Fourth, we explored only a few examples of possible methods for addressing sensitivity to MNAR data. Nonetheless, we hope our introduction of these approaches with corresponding R syntax encourages mHealth researchers to begin implementing and perhaps even testing and developing strategies for addressing the missing data realities of mHealth.

Several future directions follow naturally from this study. MNAR sensitivity analyses could be integrated into future mHealth RCTs. For instances with longitudinal data, more complex pattern-mixture models may be especially attractive [84]. For studies with fewer time points, approaches such as those introduced here may be worthwhile. If a specific sensitivity analysis approach were to become widely used, it could provide researchers with a common metric for evaluating the potential influence of differential attrition as a source of MNAR on study results. An approach based on readily interpretable metrics (eg, Cohen d) could be helpful, although there are certainly many viable possibilities, many of which may have advantages over the strategy introduced here. This is an area of active research, and new and much more sophisticated methods are regularly becoming available [76].

Short of incorporating sensitivity analyses into mHealth RCTs, researchers could, at a minimum, test for differential attrition, especially when comparing active and passive conditions. Acknowledging the potential influence of MNAR, when differential attrition is present, can allow readers to more accurately evaluate study findings in light of this limitation. A way to assess the potential impact of MNAR because of differential attrition would be through reanalysis of published mHealth RCTs, especially large trials that were seen to have higher rates of differential attrition. Reanalyses of this kind could help determine the degree to which findings are sensitive to varying MNAR assumptions, and by extension, the degree to which conclusions drawn from the broader literature may be similarly influenced. Another future direction is intentionally adopting methods that decrease attrition generally [85], given that differential attrition and associated MNAR data are less concerning when the amount of missing data is small. Finally, it could be valuable to investigate differential attrition for in-person interventions as well. To our knowledge, no such meta-analysis exists, although the same potential risk of bias because of MNAR may be applied.

Conclusions

Attrition is a persistent thorn in the side of mHealth clinical trialists [1]. Modern missing data methods such as MI and ML successfully minimize the negative impact of some types of missing data (MCAR and MAR), restoring statistical power and reducing bias in parameter estimates. However, these methods cannot remove the bias associated with MNAR data.

Evidence of differential attrition supports the possibility that MNAR may be a common problem in mHealth RCTs with passive controls and one that is largely unacknowledged to date. Sensitivity analyses offer an approach for establishing the impact of differential attrition on the study results.

Acknowledgments

The authors include data from an RCT registered at ClinicalTrials.gov (NCT04139005) and through the Open Science Framework [86]. Data are available at the Open Science Framework [39]. This research was supported by the National Center for Complementary and Integrative Health Grant K23AT010879 (SBG), the National Institute of Mental Health Grant R01MH43454 (RJD), and by the Clinical and Translational Science Award program through the National Institutes of Health National Center for Advancing Translational Sciences Grant UL1TR002373. The authors are grateful to Matthew J Hirshberg for his assistance with tidyverse code in R.

Conflicts of Interest

RJD is the founder, president, and serves on the board of directors for the nonprofit organization Healthy Minds Innovations, Inc.

Multimedia Appendix 1
R code for conducting sensitivity analysis.
[DOCX File , 29 KB - jmir_v23i6e26749_app1.docx]

Multimedia Appendix 2
Model-based approaches for handling missing not at random data.
[DOCX File , 28 KB - jmir_v23i6e26749_app2.docx]

References

1. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
2. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
3. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
4. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):20-21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
5. Bidargaddi N, Musiat P, Winsall M, Vogl G, Blake V, Quinn S, et al. Efficacy of a web-based guided recommendation service for a curated list of readily available mental health and well-being mobile apps for young people: randomized controlled trial. J Med Internet Res 2017 May 12;19(5):e141. [doi: 10.2196/jmir.6775] [Medline: 28500020]
6. Litvin S, Saunders R, Maier MA, Lüttke S. Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: a randomized controlled trial. PLoS One 2020 Sep 2;15(9):e0237220 [FREE Full text] [doi: 10.1371/journal.pone.0237220] [Medline: 32877425]
7. Fukuoka Y, Gay C, Haskell W, Arai S, Vittinghoff E. Identifying factors associated with dropout during prerandomization run-in period from an mhealth physical activity education study: the mPED trial. JMIR Mhealth Uhealth 2015 Apr 13;3(2):e34 [FREE Full text] [doi: 10.2196/mhealth.3928] [Medline: 25872754]
8. Enders CK. Applied missing data analysis. New York, United States: Guilford Press; 2010:1-377.
9. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol 2009;60:549-576. [doi: 10.1146/annurev.psych.58.110405.085530] [Medline: 18652544]
10. Blankers M, Koeter MW, Schippers GM. Missing data approaches in eHealth research: simulation study and a tutorial for nonmathematically inclined researchers. J Med Internet Res 2010;12(5):e54 [FREE Full text] [doi: 10.2196/jmir.1448] [Medline: 21169167]
11. Lüdtke T, Pult LK, Schröder J, Moritz S, Bücker L. A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. Psychiatry Res 2018 Nov;269:753-762. [doi: 10.1016/j.psychres.2018.08.113] [Medline: 30273901]
12. Bell ML, Kenward MG, Fairclough DL, Horton NJ. Differential dropout and bias in randomised controlled trials: when it matters and when it may not. Br Med J 2013 Jan 21;346:e8668 [FREE Full text] [Medline: 23338004]
13. Crutzen R, Viechtbauer W, Spigt M, Kotz D. Differential attrition in health behaviour change trials: a systematic review and meta-analysis. Psychol Health 2015 Jan;30(1):122-134. [doi: 10.1080/08870446.2014.953526] [Medline: 25109224]
14. Hedeker D, Mermelstein R, Demirtas H. Analysis of binary outcomes with missing data: missing = smoking, last observation carried forward, and a little multiple imputation. Addiction 2007 Oct;102(10):1564-1573 [FREE Full text] [doi: 10.1111/j.1360-0443.2007.01946.x] [Medline: 17854333]
15. Molenberghs G. Analyzing incomplete longitudinal clinical trial data. Biostatistics 2004 Jul 01;5(3):445-464. [doi: 10.1093/biostatistics/kxh001]
16. Heckman JJ. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In: Berg SV, editor. Annals of Economic and Social Measurement. Cambridge, Massachusetts, United States: National Bureau of Economic Research, Inc; 1972:475-492.
17. Little RJ. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc 2012 Dec 20;88(421):125-134. [doi: 10.1080/01621459.1993.10594302]
18. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002;7(2):147-177. [doi: 10.1037/1082-989x.7.2.147]
19. Nich C, Carroll KM. 'Intention-to-treat' meets 'missing data': implications of alternate strategies for analyzing clinical trials data. Drug Alcohol Depend 2002 Oct;68(2):121-130. [doi: 10.1016/S0376-8716(02)00111-4]

20. Yang X, Shoptaw S. Assessing missing data assumptions in longitudinal studies: an example using a smoking cessation trial. Drug Alcohol Depend 2005 Mar 07;77(3):213-225. [doi: 10.1016/j.drugalcdep.2004.08.018] [Medline: 15734221]
21. Cook JA, Burke-Miller J, Fitzgibbon G, Grey DD, Heflinger CA, Paulson RI, et al. Effects of alcohol and drug use on inpatient and residential treatment among youth with severe emotional disturbance in Medicaid-funded behavioral health care plans. J Psychoactive Drugs 2004 Dec;36(4):463-471. [doi: 10.1080/02791072.2004.10524429] [Medline: 15751484]
22. Baker TB, Piper ME, Stein JH, Smith SS, Bolt DM, Fraser DL, et al. Effects of nicotine patch vs varenicline vs combination nicotine replacement therapy on smoking cessation at 26 weeks: a randomized clinical trial. J Am Med Assoc 2016 Jan 26;315(4):371-379 [FREE Full text] [doi: 10.1001/jama.2015.19284] [Medline: 26813210]
23. Leurent B, Gomes M, Faria R, Morris S, Grieve R, Carpenter JR. Sensitivity analysis for not-at-random missing data in trial-based cost-effectiveness analysis: a tutorial. Pharmacoeconomics 2018 Aug 20;36(8):889-901 [FREE Full text] [doi: 10.1007/s40273-018-0650-5] [Medline: 29679317]
24. Little RJ. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc 1988 Dec;83(404):1198-1202. [doi: 10.1080/01621459.1988.10478722]
25. Byar DP, Simon RM, Friedewald WT, Schlesselman JJ, DeMets DL, Ellenberg JH, et al. Randomized clinical trials. N Engl J Med 1976 Jul 08;295(2):74-80. [doi: 10.1056/nejm197607082950204]
26. Borenstein M, Hedges L, Higgins J, Rothstein H. Introduction to Meta-analysis. New York: Wiley; 2009:1-452.
27. Goldberg SB, Imhoff-Smith T, Bolt DM, Wilson-Mendenhall CD, Dahl CJ, Davidson RJ, et al. Testing the efficacy of a multicomponent, self-guided, smartphone-based meditation app: three-armed randomized controlled trial. JMIR Ment Health 2020 Dec 27;7(11):e23825 [FREE Full text] [doi: 10.2196/23825] [Medline: 33245288]
28. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog Cardiovasc Dis 1985 Mar;27(5):335-371. [doi: 10.1016/s0033-0620(85)80003-7]
29. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. London: Wiley & Sons; 2008:1-672.
30. Sweeting M, Sutton A, Lambert P. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med 2004 May 15;23(9):1351-1375 [FREE Full text] [doi: 10.1002/sim.1761] [Medline: 15116347]
31. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
32. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Soft 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
33. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis in R: A Hands-On Guide. Boca Raton, FL and London: Chapmann & Hall/CRC Press; 2021.
34. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. URL: https://www.r-project.org/ [accessed 2021-06-01]
35. Quartagno M, Carpenter J. jomo: A package for multilevel joint modeling multiple imputation. 2020. URL: https://CRAN.R-project.org/package=jomo [accessed 2021-06-01]
36. Lumley T. mitools: Tools for multiple imputation of missing data. 2019. URL: https://CRAN.R-project.org/package=mitools [accessed 2021-06-01]
37. Buuren SV, Groothuis-Oudshoorn K. MICE: Multivariate imputation by chained equations in. J Stat Soft 2011;45(3):1-68 [FREE Full text] [doi: 10.18637/jss.v045.i03]
38. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Ed.). Hillsdale, NJ: Erlbaum; 1988.
39. Goldberg S. Missing data in mHealth RCTs. OSF Home. 2020. URL: https://osf.io/52ru3/ [accessed 2021-06-08]
40. Carissoli C, Villani D, Riva G. Does a meditation protocol supported by a mobile application help people reduce stress? Suggestions from a controlled pragmatic trial. Cyberpsychol Behav Soc Netw 2015 Jan;18(1):46-53. [doi: 10.1089/cyber.2014.0062] [Medline: 25584730]
41. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
42. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
43. 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 Mar;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
44. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. PLoS One 2018 Dec 31;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]
45. Enock PM, Hofmann SG, McNally RJ. Attention bias modification training via smartphone to reduce social anxiety: a randomized, controlled multi-session experiment. Cogn Ther Res 2014 Mar 4;38(2):200-216. [doi: 10.1007/s10608-014-9606-z]
46. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as an electronic biomarker of illness activity in bipolar disorder. Bipolar Disord 2015 Nov;17(7):715-728. [doi: 10.1111/bdi.12332] [Medline: 26395972]

47. Hall BJ, Xiong P, Guo X, Sou EK, Chou UI, Shen Z. An evaluation of a low intensity mHealth enhanced mindfulness intervention for Chinese university students: a randomized controlled trial. Psychiatry Res 2018 Dec;270:394-403. [doi: 10.1016/j.psychres.2018.09.060] [Medline: 30300870]
48. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile phone-delivered cognitive behavioral therapy for insomnia: a randomized waitlist controlled trial. J Med Internet Res 2017 Apr 11;19(4):e70 [FREE Full text] [doi: 10.2196/jmir.6524] [Medline: 28400355]
49. Ivanova E, Lindner P, Ly KH, Dahlin M, Vernmark K, Andersson G, et al. Guided and unguided acceptance and commitment therapy for social anxiety disorder and/or panic disorder provided via the internet and a smartphone application: a randomized controlled trial. J Anxiety Disord 2016 Dec;44:27-35. [doi: 10.1016/j.janxdis.2016.09.012] [Medline: 27721123]
50. Kahn JR, Collinge W, Soltysik R. Post-9/11 veterans and their partners improve mental health outcomes with a self-directed mobile and web-based wellness training program: a randomized controlled trial. J Med Internet Res 2016 Sep 27;18(9):e255 [FREE Full text] [doi: 10.2196/jmir.5800] [Medline: 27678169]
51. Krafft J, Potts S, Schoendorff B, Levin ME. A randomized controlled trial of multiple versions of an acceptance and commitment therapy matrix app for well-being. Behav Modif 2019 Mar;43(2):246-272. [doi: 10.1177/0145445517748561] [Medline: 29262693]
52. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain. part 2: 11-month follow-up results of a randomized trial. J Med Internet Res 2013 Mar 28;15(3):e72 [FREE Full text] [doi: 10.2196/jmir.2442] [Medline: 23538392]
53. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
54. Lee RA, Jung ME. Evaluation of an mHealth app (DeStressify) on university students' mental health: pilot trial. JMIR Ment Health 2018 Jan 23;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]
55. Levin ME, Haeger J, Pierce B, Cruz RA. Evaluating an adjunctive mobile app to enhance psychological flexibility in acceptance and commitment therapy. Behav Modif 2017 Nov;41(6):846-867. [doi: 10.1177/0145445517719661] [Medline: 28689449]
56. Levin ME, Haeger J, An W, Twohig MP. Comparing cognitive defusion and cognitive restructuring delivered through a mobile app for individuals high in self-criticism. Cogn Ther Res 2018 Jul 13;42(6):844-855. [doi: 10.1007/s10608-018-9944-3]
57. Lukas CA, Berking M. Reducing procrastination using a smartphone-based treatment program: a randomized controlled pilot study. Internet Interv 2018 Jun;12:83-90 [FREE Full text] [doi: 10.1016/j.invent.2017.07.002] [Medline: 30135772]
58. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
59. Marx LS. A mindful eating “app” for non-treatment-seeking university women with eating and weight concerns. Ph D Dissertation - Emory University, Georgia, United States. 2016. URL: https://search.proquest.com/docview/1896581612/abstract/8A3BA70DAC8C4214PQ/1 [accessed 2020-12-07]
60. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PSTD coach app: a pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 Jan 25;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
61. Moëll B, Kollberg L, Nasri B, Lindefors N, Kaldo V. Living SMART — A randomized controlled trial of a guided online course teaching adults with ADHD or sub-clinical ADHD to use smartphones to structure their everyday life. Internet Interv 2015 Mar;2(1):24-31. [doi: 10.1016/j.invent.2014.11.004]
62. Oh SJ, Seo S, Lee JH, Song MJ, Shin M. Effects of smartphone-based memory training for older adults with subjective memory complaints: a randomized controlled trial. Aging Ment Health 2018 Apr;22(4):526-534. [doi: 10.1080/13607863.2016.1274373] [Medline: 28071929]
63. Pham Q, Khatib Y, Stansfeld S, Fox S, Green T. Feasibility and efficacy of an mhealth game for managing anxiety: "Flowy" randomized controlled pilot trial and design evaluation. Games Health J 2016 Feb;5(1):50-67. [doi: 10.1089/g4h.2015.0033] [Medline: 26536488]
64. Proudfoot J. The future is in our hands: the role of mobile phones in the prevention and management of mental disorders. Aust N Z J Psychiatry 2013 Feb;47(2):111-113. [doi: 10.1177/0004867412471441] [Medline: 23382507]
65. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of superbetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
66. Rosen KD, Paniagua SM, Kazanis W, Jones S, Potter JS. Quality of life among women diagnosed with breast cancer: a randomized waitlist controlled trial of commercially available mobile app-delivered mindfulness training. Psychooncology 2018 Aug;27(8):2023-2030. [doi: 10.1002/pon.4764] [Medline: 29766596]

67. Schlosser D, Campellone T, Truong B, Etter K, Vergani S, Komaiko K, et al. Efficacy of PRIME, a mobile app intervention designed to improve motivation in young people with schizophrenia. Schizophr Bull 2018 Aug 20;44(5):1010-1020 [FREE Full text] [doi: 10.1093/schbul/sby078] [Medline: 29939367]
68. Stjernswärd S, Hansson L. Effectiveness and usability of a web-based mindfulness intervention for caregivers of people with mental or somatic illness. A randomized controlled trial. Internet Interv 2018 Jun;12:46-56 [FREE Full text] [doi: 10.1016/j.invent.2018.03.004] [Medline: 30135768]
69. Stolz T, Schulz A, Krieger T, Vincent A, Urech A, Moser C, et al. A mobile app for social anxiety disorder: a three-arm randomized controlled trial comparing mobile and PC-based guided self-help interventions. J Consult Clin Psychol 2018 Jun;86(6):493-504. [doi: 10.1037/ccp0000301] [Medline: 29781648]
70. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Dec 27;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
71. van Emmerik AA, Berings F, Lancee J. Efficacy of a mindfulness-based mobile application: a randomized waiting-list controlled trial. Mindfulness (N Y) 2018;9(1):187-198 [FREE Full text] [doi: 10.1007/s12671-017-0761-7] [Medline: 29387266]
72. Versluis A, Verkuil B, Spinhoven P, Brosschot J. Effectiveness of a smartphone-based worry-reduction training for stress reduction: a randomized-controlled trial. Psychol Health 2018 Sep;33(9):1079-1099. [doi: 10.1080/08870446.2018.1456660] [Medline: 29609475]
73. Yang E, Schamber E, Meyer RML, Gold JI. Happier Healers: randomized controlled trial of mobile mindfulness for stress management. J Altern Complement Med 2018 May;24(5):505-513. [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
74. Allison PD. Missing Data. Thousand Oaks, California, United States: SAGE Publications; 2001:1-104.
75. Manning W, Duan N, Rogers W. Monte Carlo evidence on the choice between sample selection and two-part models. J Econom 1987 May;35(1):59-82. [doi: 10.1016/0304-4076(87)90081-9]
76. Xie H, Gao W, Xing B, Heitjan DF, Hedeker D, Yuan C. Measuring the impact of nonignorable missingness using the R package isni. Comput Methods Programs Biomed 2018 Oct;164:207-220 [FREE Full text] [doi: 10.1016/j.cmpb.2018.06.014] [Medline: 30195428]
77. Valentine JC, Pigott TD, Rothstein HR. How many studies do you need? J Educ Behav Stat 2010 Apr 01;35(2):215-247. [doi: 10.3102/1076998609346961]
78. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
79. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
80. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
81. Heber E, Ebert DD, 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]
82. Simon GE, Imel ZE, Ludman EJ, Steinfeld BJ. Is dropout after a first psychotherapy visit always a bad outcome? Psychiatr Serv 2012 Jul;63(7):705-707 [FREE Full text] [doi: 10.1176/appi.ps.201100309] [Medline: 22752034]
83. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3(1):21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
84. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
85. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med 2012 Oct 4;367(14):1355-1360 [FREE Full text] [doi: 10.1056/NEJMsr1203730] [Medline: 23034025]
86. Impact of targeted mental exercises on awareness, connection, and insight: a pilot RCT. Open Science Framework Registries. URL: https://osf.io/rdey9 [accessed 2021-06-09]

Abbreviations

MAR: missing at random
MCAR: missing completely at random
mHealth: mobile health

MI: multiple imputation
ML: maximum likelihood
MNAR: missing not at random
OR: odds ratio
RCT: randomized controlled trial

Edited by R Kukafka; submitted 23.12.20; peer-reviewed by M Blankers, O Bur, S Shoop-Worrall; comments to author 21.01.21; revised version received 01.02.21; accepted 06.05.21; published 15.06.21.

Please cite as:
Goldberg SB, Bolt DM, Davidson RJ
Data Missing Not at Random in Mobile Health Research: Assessment of the Problem and a Case for Sensitivity Analyses
J Med Internet Res 2021;23(6):e26749
URL: https://www.jmir.org/2021/6/e26749
doi:10.2196/26749
PMID:34128810

Original Paper

Effects of Self-focused Augmented Reality on Health Perceptions During the COVID-19 Pandemic: A Web-Based Between-Subject Experiment

Ayanna Seals[1], MSc; Monsurat Olaosebikan[2], MSc; Jennifer Otiono[3], BA; Orit Shaer[3], PhD; Oded Nov[4], PhD

[1]New York University, Brooklyn, NY, United States
[2]Tufts University, Medford, MA, United States
[3]Wellesley College, Wellesley, MA, United States
[4]Department of Technology Management and Innovation, Tandon School of Engineering, New York University, Brooklyn, NY, United States

Corresponding Author:
Ayanna Seals, MSc
New York University
5 Metrotech Center, 4th Fl
Brooklyn, NY, 11201
United States
Phone: 1 646 997 3760
Email: ayannaseals@nyu.edu

Abstract

Background: Self-focused augmented reality (AR) technologies are growing in popularity and present an opportunity to address health communication and behavior change challenges.

Objective: We aimed to examine the impact of self-focused AR and vicarious reinforcement on psychological predictors of behavior change during the COVID-19 pandemic. In addition, our study included measures of fear and message minimization to assess potential adverse reactions to the design interventions.

Methods: A between-subjects web-based experiment was conducted to compare the health perceptions of participants in self-focused AR and vicarious reinforcement design conditions to those in a control condition. Participants were randomly assigned to the control group or to an intervention condition (ie, self-focused AR, reinforcement, self-focus AR × reinforcement, and avatar).

Results: A total of 335 participants were included in the analysis. We found that participants who experienced self-focused AR and vicarious reinforcement scored higher in perceived threat severity (*P*=.03) and susceptibility (*P*=.01) when compared to the control. A significant indirect effect of self-focused AR and vicarious reinforcement on intention was found with perceived threat severity as a mediator (b=.06, 95% CI 0.02-0.12, SE .02). Self-focused AR and vicarious reinforcement did not result in higher levels of fear (*P*=.32) or message minimization (*P*=.42) when compared to the control.

Conclusions: Augmenting one's reflection with vicarious reinforcement may be an effective strategy for health communication designers. While our study's results did not show adverse effects in regard to fear and message minimization, utilization of self-focused AR as a health communication strategy should be done with care due to the possible adverse effects of heightened levels of fear.

(J Med Internet Res 2021;23(6):e26963) doi:10.2196/26963

KEYWORDS

COVID-19; health behavior; augmented reality; self-focused attention; vicarious reinforcement; human-computer interactions; hand hygiene; perception

Introduction

Overview

With self-focused augmented reality (AR) usage increasing in recent years [1], the utilization of this technology has a potential in addressing health communication and behavior interventions challenges. Whereas AR technology layers digital content onto the real world [2], self-focused AR visually augments the self, layering digital content onto the self. One technology enabling self-focused AR is video filters, which superimpose computer-generated content onto a user using their web or smartphone camera (eg, Snapchat Lenses [3] and AR effects on Instagram [4]). For example, Snapchat's Time Machine AR lens morphs a user's reflection to display what the user might look like at different ages. Smart mirrors, another self-focused AR technology, combine digital screens (eg, an LCD [liquid-crystal display] monitor) with semitransparent glass. A smart mirror looks and functions similarly to a traditional mirror but with digital content displayed in the foreground [5]. Real-world applications of smart mirrors are on display in vehicles [6], dressing rooms [7], and home gyms [8]. For example, the MIRROR home gym [8] displays fitness instruction layered on top of the user's reflection. Both AR video filters and smart mirrors provide new opportunities for displaying health behavior communication to the public.

In response to the recent increase in self-focused AR usage [1,6-8], we investigated the potential impact of self-focused AR within the health domain. For behavior change researchers, the effect of layering health threats along with mitigative behaviors and their results on top of a user may be of particular interest. For example, would using AR to layer a set of very healthy teeth due to good oral hygiene on dental patients impact their behavior? Could using AR in rearview mirrors to overlay scratches and bruises on top of drivers and riders encourage seat belt usage? We now have self-focused AR technologies in the hands of millions, presenting the opportunity to visually show individuals the impact of their decisions before they make them.

Research in psychology suggests that heightening self-focused attention (manipulated by using a mirror or video camera) has implications for perception, affective experiences (emotions, feelings, and moods), and behavior [9]. Objective self-awareness occurs when an individual places attention on themself, viewing themself as a social object. Objective self-awareness theory [10] posits that self-focused attention heightens the awareness of the gap between one's perceived "real self" and "ideal self," resulting in negative affect. For example, if one desires good health and believes that exercise is vital to maintain one's health and yet does not exercise, heightening objective self-awareness will likely result in negative emotions. The increased negative affect resulting from the awareness of discrepancies leads to either (1) the avoidance of self-focused attention and the discrepancy or (2) actions to reduce the discrepancy [10]. This theory suggests that self-focused AR might impact behavior. We investigated which perceptions could be involved when individuals experience self-focused AR within a health context. Research prototypes have explored self-focus [11] and self-focused AR [5,12-14] technologies for health behavior change. However, few have investigated how the design of interventions that aim to increase self-focused attention might impact health perceptions and emotions. Similarly, prior studies did not consider the potential of combining self-focused AR with vicarious reinforcement, that is, reinforcement from observing others' behavior and the results of those actions. This paper draws upon insights from objective self-awareness theory [2] and social cognitive theory [15] to inform hypotheses about the relationships between predictors of health behavior change and self-focused AR.

We present findings from an online experiment on the impact of combining self-focused AR with vicarious reinforcement, visualizing the cause and effect of risk-mitigating behavior layered onto one's reflection. Our study took place during the COVID-19 pandemic, focusing on hand hygiene behavior as an effective measure against pathogen transmission [16,17]. We discuss the implications of the results in light of the public health emergency, addressing the following research question: how does reinforcement in self-focused AR impact health perceptions during a pandemic?

Background

Various health behavior change models [18,19] highlight the roles of predictors of intentions such as risk perceptions (perceived threat severity and threat susceptibility) and outcome expectancy. Drawing on objective self-awareness theory and social cognitive theory, we postulated that health behavior change–themed self-focused AR could impact these predictors of intention.

Research suggests that self-focused attention can result in action consistency with behavioral standards [9]. Objective self-awareness theory posits that self-focused attention will result in negative affect through the increased awareness of contradicting beliefs about one's self or discrepancies between belief and behavior [10]. If negative affect is experienced, and one does not avoid the self-focused attention, they will attempt to reduce the discrepancy to reduce the negative affect, such as by changing their behavior. To further illustrate this, recall the individual from the example earlier who values their health and believes that exercise is vital for their health but does not exercise. According to objective self-awareness theory, an increase in self-focused attention would result in an attempt to reduce the discrepancy, which may result in exercise behavior. We propose that when increasing self-focused attention in the context of a health threat, especially during a pandemic, that the negative affect experienced will include fear and will increase to levels higher than if self-focused attention was not activated.

Research suggests that fear may play a large role in health behavior, especially during public health emergencies such as the current pandemic. Harper et al [20] found COVID-19 fear scores to be a positive predictor of behavior change. Fear may also impact behavior as it relates to risk perception. Risk perception, an individual's perceived susceptibility to or severity of a threat, is included in many health behavior change theories [21]. Li [22] found perceived threat (measured by averaging threat susceptibility and severity) to be a positive predictor of fear. Affective factors are believed to play a role in the formation of risk perception [23]. While risk perceptions can increase fear, fear has also been found to induce higher risk perceptions [24]. Self-focused AR content layering onto the body increases the sense of spatial presence (ie, the object "being there"), potentially heightening fear if the object is threatening. Due to the combination of self-focus and spatial presence, we suggest that health threat–related self-focused AR may impact levels of fear, perceived threat severity, and perceived threat susceptibility.

It is important to consider the potential negative impact of heightening fear and risk perceptions. While Harper et al [20] found increased fear to be associated with higher behavioral adherence, they also found fear to be correlated with decreased physical and environmental quality of life and warned about mental health implications. Fear can also have an adverse effect on behavior. The Extended Parallel Process Model (EPPM) [25] outlines the importance of a balance of fear and efficacy for health communication campaigns to be effective. While fear can be a motivator for behavior, where the fear/efficacy balance is disrupted, individuals may use cognitive defense mechanisms instead of behavior as a means of fear control. In this case, not only would the behavior change method be ineffective, but it could result in the adverse effect of prompting the development of these defense mechanisms. Based on the EPPM, Li [22] tested a model for protective behaviors during a public health emergency with a study during the Ebola outbreak of 2014. Fear controls measured included negative reactance to messages, message minimizing, and defensive avoidance. Li [22] found perceived threat to have a significant effect on fear and fear controls but did not find self-efficacy to be a successful moderator of that relationship. This suggests that although fear may be an effective strategy to encourage health behavior adherence, certain levels of fear may lead individuals to minimize health behavior messaging to control their fear instead of engaging in behavior change. We take these findings and the EPPM into consideration, as our study directly layers a health threat onto participants, which could result in excessive levels of fear triggering adverse fear control mechanisms. We expected self-focused AR, in the form of AR video filters, to heighten both fear and fear control mechanisms when displaying a health threat.

Our study investigated the impact of vicarious reinforcement outcome expectancy when combined with self-focused AR. Research suggests that outcome expectancy mediates the impact of self-focused attention on behavior. For someone who has been made aware of a discrepancy between "actual state" and "desired state," if they don't believe a suggested behavior change will result in the "desire state," they are more likely to change the "desired state" [9]. When the "desired state" is health related, this can have adverse implications. Outcome expectancy can be impacted by experiencing vicarious reinforcement. Vicarious reinforcement occurs when a reinforcing effect for an individual takes place by observing others' behavior and the results of their actions [26]. Bandura et al [27] found that children who were exposed to media displaying aggressive behavior that was rewarded showed more imitative aggressive behavior than those who saw aggressive behavior that was punished. Bandura's [15] social cognitive theory details how behaviors can be formed by observing a model engage in a behavior. Bandura suggests that due to limited contact with physical and social environments, people rely largely on vicarious experiences to form their idea of reality. In our study, vicarious reinforcement consists of visual representations of pathogens (ie, germs), which are made visible on avatar hands. As a hand hygiene animation plays, covering all the steps of proper handwashing, these pathogens disappear from the avatar's hands. We predicted that this experience would heighten one's perception of outcome expectancy. Given that health behavior models, such as The Health Action Process Approach [18] attribute outcome expectancies to the formation of intention, we found this valuable to investigate in our study. We proposed that the vicarious experience described above would directly affect outcome expectancy.

Related Work

While studies combining self-focused attention and vicarious reinforcement have yet to see much direct utilization in human-computer interaction research, a few studies on health smart mirrors [5,12,13], self-representation [28], and spatial presence [14] suggest an impact of self-focused attention on perceptions and behavior.

While applications of smart mirrors for health care are limited, exploratory research prototypes have shown their potential to detect emotional states, monitor physiological parameters, and encourage behavior change. The Wize Mirror [12] encouraged users to improve their lifestyle to mitigate cardiometabolic risk assessed by tracking physical face signs)eg, skin color, subcutaneous fat, facial expressions). Medical Mirror [5] utilized computer vision and advanced signal processing within a smart mirror design to encourage people to keep track of their vital signs regularly. Fit Mirror [13] increased user's motivation, happiness, and fitness for the day by integrating exercising and challenging others into their morning routine. Although the rise of smart mirrors has resulted in studies exploring the use of these devices in health care, there is a lack of research investigating how self-focus specifically plays a role in influencing health behavior change. The studies mentioned above lack a control condition in which all design features are present except the mirror to study the direct impact of seeing one's self-reflection.

A recent study by Jung et al [14] used a projector and a mirror to show participants their bodies with x-ray visualization of smoking lungs. A separate condition displayed the same content but on a mannequin. Both conditions were compared to a control, which displayed the information on a screen in 2D. The researchers found that spatial augmented reality increases spatial presence, the perceptual illusion that the real world and the mediated world are "equally present." In addition, they found that higher levels of spatial presence were associated with a negative emotional change toward cigarettes and cigarette cessation campaign engagement intention. These findings support the idea that displaying the consequences of health behaviors on top of the user's own body can impact behavior change constructs regarding emotions and intention. Our study aimed to contribute to this line of research by exploring additional behavior change constructs.

One notable finding from Jung et al [14] was that the mannequin condition also resulted in higher levels of spatial presence. However, an analysis comparing emotions and intentions reported for those in the mannequin condition compared to the control was not reported. Gaining a better understanding of how objects that can serve as external self-representations, such as avatars and mannequins, could help develop more feasible design interventions when mapping AR elements directly on the body would be complex.

Yee and Bailenson [28] found that self-representations can help form our behaviors, even when this representation is digital, such as in the case of avatars. Yee and Bailenson [28] call this the Proteus effect and provided support from two experiments. In study one, participants were provided with an avatar that was previously rated as high, medium, or low on an attractiveness scale. They were asked to interact with another character (in a virtual reality environment) after looking at themselves in a mirror. Those in the high attractiveness condition disclosed more information and moved closer to the other character. The second experiment, testing avatar height in an ultimatum game, found that those in the tall condition were more likely to offer an unfair split. Those in the short condition were more likely to accept an unfair split. These findings suggest that augmentations to self-representation, as an avatar, may impact one's behavior. Fox and Bailenson [29] studied whether vicarious reinforcement with a user's avatar had an effect on physical exercise. Seeing one's avatar benefit from exercise behavior and experience consequences from not engaging in the behavior encouraged the observer to engage in the behavior. These results suggest that vicarious reinforcement using avatars may be effective. Based on the studies mentioned above, we expected that the display of health threats on an avatar representation of the self in an AR environment will impact levels of threat severity, susceptibility, fear, and message minimization.

Our study expands on existing research by investigating how health behavior self-focused AR may impact specific predictors of behavioral intentions and what negative implications may exist in regard to fear control responses.

This Study

In our research, we examined the impact of self-focused AR and vicarious reinforcement on perception and emotion as it relates to hand washing health beliefs and behavioral intentions. Below, we present our hypotheses:

> *H1: The combination of self-focused AR and vicarious reinforcement will result in higher levels of perceived positive outcome expectancy, perceived threat severity and susceptibility, fear, and message minimization when compared to a control.*
>
> *H2: Using avatar representations in self-focused AR with vicarious reinforcement will result in higher levels of threat severity, susceptibility, fear, and message minimization compared to a control.*

Methods

To study the effects of self-focused AR on behavioral intention and perception, we conducted an online experiment. Participants interacted (Figure 1) with a web application (Figure 2) that displayed health information regarding the coronavirus and a hand hygiene animation (Figure 3). Five conditions differed in their inclusion of self-focused attention and vicarious reinforcement.

Figure 1. A participant in one of the self-focused augmented reality design groups viewing a reflection of themselves using a video feed from their camera.

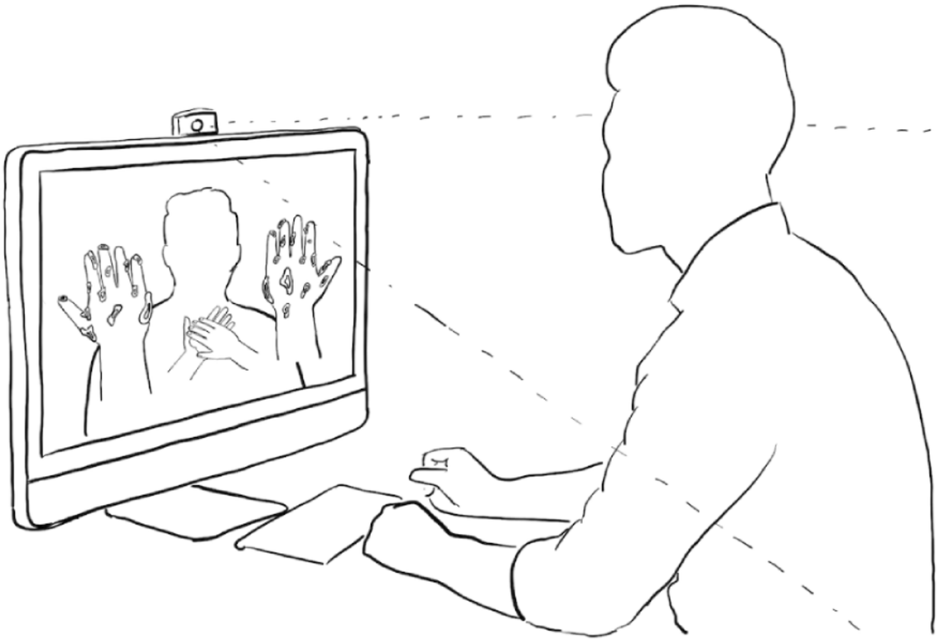


Figure 2. A screenshot from the web application.

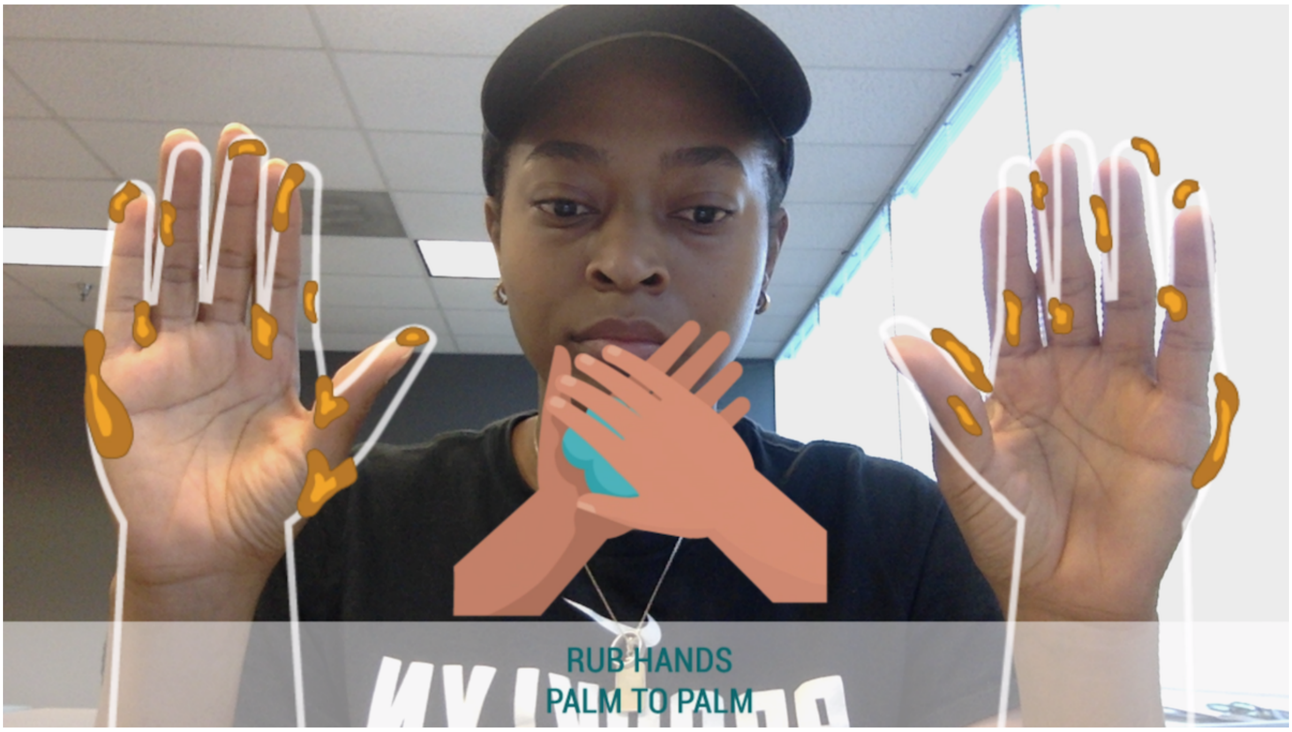


Figure 3. An example of the progression of vectors used in the handwashing animation.

Experimental Conditions

We conducted a between-subjects experiment where users interacted with a web application and then responded to an online questionnaire. The study followed a posttest-only control group design to avoid a testing threat to internal validity. The design of the web application differed depending on the intervention condition each participant was randomly allocated to. All five conditions displayed the same information about COVID-19, including how it is spread and preventative measures as described by the US Centers for Disease Control and Prevention (CDC) [17] and the World Health Organization (WHO) [30]. This information was followed by a hand hygiene animation, the display of which differed based on the participant's assigned condition.

Control Condition: No Self-focused AR or Vicarious Reinforcement

In the control condition (Figure 4), we displayed an animation of a 12-step handwashing technique following standards outlined by the WHO [30], accompanied by captions to describe each movement ("Rub hands palm to palm").

Figure 4. Control condition.

Reinforcement

In the reinforcement condition (Figure 5), the handwashing animation described in the control condition was accompanied by an additional animation showing germs disappearing from a pair of illustrated hands as the handwashing animation progressed. These animations were synced so that the appropriate areas of the hands displayed were affected based on the specific stage of the handwashing animation a viewer was watching. For example, the thumb cleaning animation segment was paired with germs disappearing from the thumbs.

Figure 5. Reinforcement condition.

Self-focused AR

The self-focus condition (Figure 6) utilized the participant's web camera to display their self-reflection, serving as the stimulus for self-focused attention. The handwashing animation was layered on top of the viewer's reflection. This reflection was shown in real time and was created using the participant's web camera.

Figure 6. Self-focused augmented reality condition.

Self-focus AR × Reinforcement

The self-focus AR × reinforcement condition (Figure 7) visualized germs directly on the participant's hands. Instructions at the beginning of the animation directed participants on where to place their hands. The handwashing animation was displayed in between their hands. As the animation progresses, the user saw the germs disappear from the reflection of their own hands.

Figure 7. Self-focus augmented reality × reinforcement condition.

Avatar

In the avatar condition (Figure 8), participants viewed an animation showing germs disappearing from a pair of illustrated hands layered on top of the user's reflection. These are referred to as avatar hands, as they are meant to represent the user's hands. The perspective displayed was that which is seen if the individual were to hold up their hands and look at them.

Figure 8. Avatar condition.

Participants

Participants were recruited via Prolific (a crowdsourcing platform [31]) and were compensated $2.50 for their time. Participation was limited to those residing in the United States, who spoke fluent English, and were ≥18 years of age. Pilot testing revealed technical challenges that could interfere with the study, mainly involving web camera use verification. To address this, prior to being recruited for the study, Prolific members were invited to a prescreener that verified (1) their access to an acceptable browser for the study (Safari, Chrome, and Firefox) and (2) that camera permissions worked with their technical setup. Prolific IDs for those who passed the screener were collected, and access to the main study was restricted to those IDs.

Procedure

The experiment took place between August 6-21, 2020. In all five conditions, after receiving consent, we described the experiment as a study on health information presentation and provided instructions to review the information given carefully. In the three self-focused-AR–based conditions, we displayed information on how to set up the web camera for the study.

All conditions provided information about COVID-19. Details focused on how the virus is spread and preventative measures as described by the CDC [17]. Participants next viewed an animation detailing the steps of proper hand hygiene as described by the WHO [30]. This was followed by a questionnaire to collect demographic information and measure health knowledge and perception, which concluded the study.

The questionnaire included questions to check whether the participant is paying attention. Three multiple-choice questions asked the user about the information displayed in the study (eg, What is a recommended preventative measure to reduce the spread of the coronavirus?). To validate that self-focused AR interventions were delivered correctly using the camera, participants in these conditions were informed, prior to the study, that screenshots would be collected randomly throughout the animation. The screenshots were reviewed to ensure that participant’s reflections were displayed to them and that those in the self-focus AR × reinforcement condition had their hands within the view as instructed. Only those who followed the instructions, verified by screenshots, were included in the final data set.

Variables and Measures

We collected measures of self-reported health beliefs, behavioral intention, and self-reported perceptions of COVID-19 (Table 1), along with demographic data. These measures were adapted from Schwarzer [18] and Li [22]. Items in this study were all measured on a 7-point Likert-type scale ranging from 1 (strongly disagree) to 7 (strongly agree).

Although adapted from previous research, cross loading was a concern due to the rewording of items and the difference in factors present compared to the adapted questionnaires. For example, we added items to measure opinions about perceived threat severity and susceptibility of family and friends. To examine the validity and reliability of our measures, we conducted exploratory and confirmatory factor analysis using a split-sample approach, with one half to develop a model and the other half to validate. Factors loaded as expected (Multimedia Appendix 1).

Data on demographics and the COVID-19 risk of a severe illness of loved ones were collected. Participants were asked to report their age, gender, and the state in which they currently reside. In addition, they were asked to report if they have a family member or friend who is at high risk of severe illness if they are infected with COVID-19. It was noted that one is considered high risk if they are ≥65 years and/or have underlying medical conditions.

Table 1. Questionnaire items.

Variable and code	Questionnaire item
Intention	
inte	I intend to wash my hands, as instructed in this study, *on a regular basis.*
Perceived outcome expectancy	
expe1	I believe proper handwashing, as instructed in this study, will help make me *less likely* to get the coronavirus disease (COVID-19).
expe2	I believe proper handwashing, as instructed in this study, will help *reduce the spread* of the coronavirus disease (COVID-19).
Fear	
fear1	The emotion that I am feeling about the coronavirus (COVID-19) pandemic is: …Frightened
fear2	…Scared
fear3	…Anxious
Message minimization	
reac1	To what extent do you feel that *preventative measures messaging*, in your state, regarding the coronavirus disease (COVID-19) is: …Manipulative
reac2	…Misleading
reac3	…Distorted
Perceived threat-severity	
seve1	I believe that the coronavirus disease (COVID-19) is a serious threat to my personal health.
seve2	I believe that the coronavirus disease (COVID-19) is a serious threat to my family members (immediate or extended).
seve3	I believe that the coronavirus disease (COVID-19) is a serious threat to my friends.
seve4	I believe that the coronavirus disease (COVID-19) is a serious threat to the general public.
Perceived threat-susceptibility	
susc1	I am at risk of catching the coronavirus disease (COVID-19).
susc2	My family (immediate or extended) members are at risk of catching the coronavirus disease (COVID-19).
susc3	My friends are at risk of catching the coronavirus disease (COVID-19).

Statistical Analysis

Analysis of our data using histograms and the Shapiro-Wilk test showed that the data were not normally distributed. Shapiro-Wilk *P* values ranged from 6.35e-08 (efficacy) to 1.927e-25 (intention). Therefore, hypothesis testing was conducted using the nonparametric Wilcoxon-Mann-Whitney test, comparing perceived threat severity, susceptibility, outcome expectancy, fear, and message minimization scores between intervention conditions. If condition pairs had the same distribution shape, medians were compared. If the shapes were different, the mean ranks were compared. Additionally, mediation models for dependent variables and design conditions with significant findings were tested. A bootstrapping method using PROCESS macro models 4 and 6 [32], 5000 bootstrap samples, and percentile bootstrap CIs were used. Significance was established at *P*<.05. Statistical analysis was performed using Python (Python Software Foundation) [33], and the pandas (Community) [34] and SciPy (Enthought) libraries [35] were used to conduct the Wilcoxon-Mann-Whitney test. SPSS software (IBM Corp) [36] was resourced, in which the PROCESS macro [32] was implemented to test mediation.

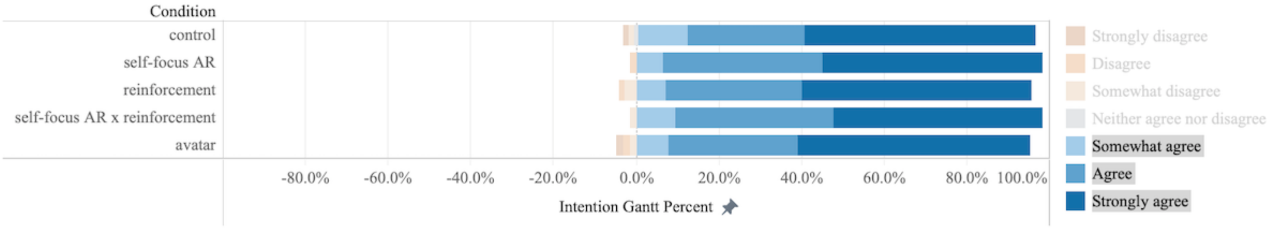
Results

Overview

A total of 502 individuals participated in the study. Of this, 335 participants met the attention and screenshot verification checks (see *Procedure* section) and were included in the analysis. Of the 335 participants, 77 were randomly assigned to the control condition, 61 to the self-focused AR condition, 70 to the reinforcement condition, 63 to the self-focus AR × reinforcement condition, and 64 to the avatar condition.

Although our study focused on predictors of behavioral intentions, we began with results pertaining to intention (Figure 9) to provide context for further discussion. A significant difference between design conditions and the control was not found (self-focused AR: *P*=.42; reinforcement: *P*=.43; self-focus AR × reinforcement: *P*=.41; avatar: *P*=.43).

Figure 9. Responses for behavioral intention. Responses are strongly skewed toward higher levels of agreement for all conditions, indicating a ceiling effect. AR: augmented reality.



Effects of Self-focused AR With Vicarious Reinforcement in Regards to Perceived Fear, Threat Severity, Threat Susceptibility, Outcome Expectancy, and Message Minimization (Hypothesis 1)

Hypothesis 1 proposed that a combination of vicarious reinforcement and self-focused AR would result in higher levels of fear, perceived threat severity, susceptibility, outcome expectancy, and message minimization compared to the control group. Compared with participants in the control (P=.43), message minimization scores of those in the self-focus AR × reinforcement condition were not significantly different. This was also the case for outcome expectancy (P=.41) and fear (P=.23) (Figure 10). However, perceived threat severity and susceptibility had significant findings.

For perceived threat severity, median scores for the self-focus AR × reinforcement and control groups were 6.25 and 6.00, respectively (Figure 11); the two groups' distributions differed significantly (Mann-Whitney U=1983, P=.03). Regarding perceived threat susceptibility, median scores for the self-focus AR × reinforcement and control groups were 6.00 and 5.33 (Figure 12); the two groups' distributions differed significantly (Mann-Whitney U=1897.0, P=.01). Our results partially supported H1 regarding perceived threat severity and susceptibility; however, we did not find support for outcome expectancy, fear, and message minimization.

Figure 10. Responses for outcome expectancy. Responses are strongly skewed toward higher levels of agreement for all conditions, indicating a ceiling effect. AR: augmented reality.

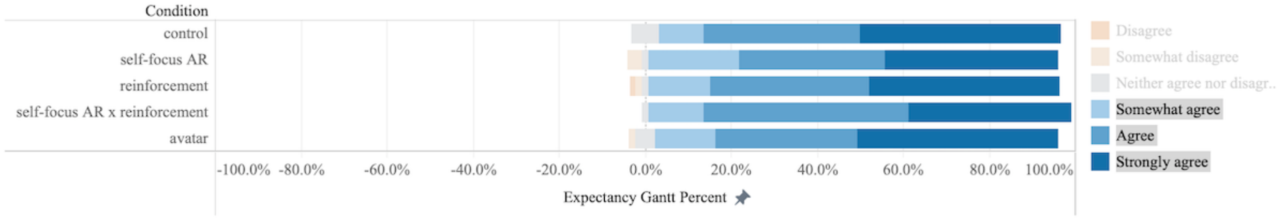


Figure 11. Significantly higher levels of perceived threat severity among participants in the self-focus augmented reality (AR) × reinforcement condition compared to the control condition. No significant differences were found between the control group and the other conditions.

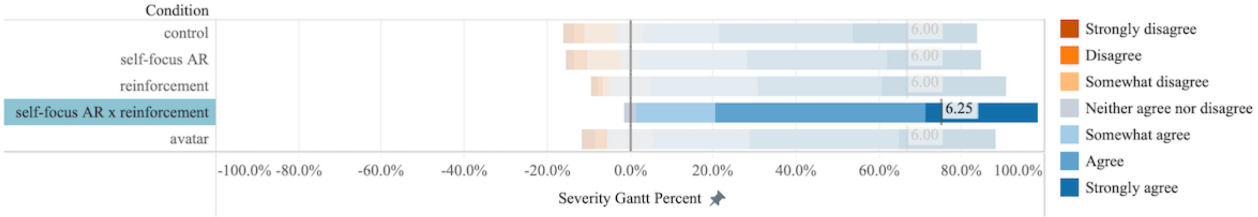
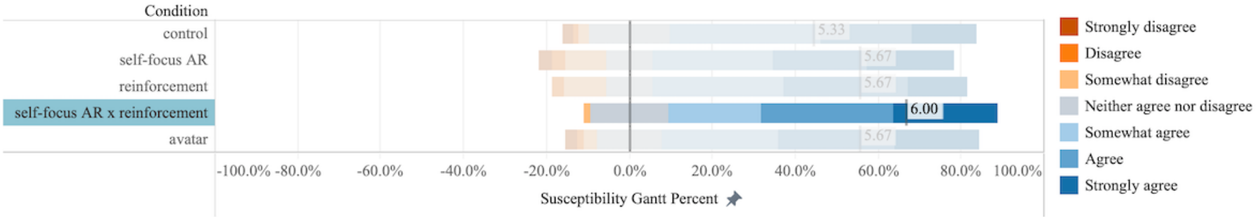


Figure 12. Significantly higher levels of perceived threat susceptibility among participants in the self-focus augmented reality (AR) × reinforcement condition compared to the control condition. No significant differences were found between the control group and the other conditions.



Additional analysis revealed that the use of self-focused AR and vicarious reinforcement individually did not impact measured predictors of intention, except for in the case of fear. When compared with participants in the control condition, those in the reinforcement condition (without self-focused AR) did not have a significant difference in medians for outcome expectancy (P=.48), threat severity (P=.39), susceptibility (P=.40), fear (P=.10), and message minimization (P=.47). Conversely, when compared to the control, those in the self-focused AR condition (without vicarious reinforcement) did not have a significant difference in medians for outcome expectancy (*P*=.26), threat severity (*P*=.21), susceptibility (*P*=.45), and message minimization (*P*=.39). Median scores for the self-focused AR and control groups were 5.0 and 5.3 (*U*=1950.5, *P*=.04).

A mediation model was used to test whether self-focus AR × reinforcement affects behavioral intention through perceived threat susceptibility and perceived threat severity (Figure 13). We found a significant indirect effect of self-focus AR × reinforcement on intention with perceived threat severity as the only mediator (b=.06, 95% CI 0.02-0.12, SE 0.02), but not with perceived threat susceptibility as the only mediator. In addition, a significant indirect effect of self-focus AR × reinforcement on intention was found when both perceived threat susceptibility and threat severity were included as serial mediators (b=.06, 95% CI 0.02-0.12, SE 0.03). These results indicate that although the self-focus AR × reinforcement condition does not directly affect intention in this study, its effect on threat susceptibility and threat severity results in an indirect effect on intention. A separate mediation model was used to investigate whether self-focus AR × reinforcement affects perceived threat severity through perceived threat susceptibility. A significant indirect effect of self-focus AR × reinforcement on perceived threat severity through perceived threat susceptibility was found (b=.26, 95% CI 0.08-0.45, SE 0.09).

Although fear was not significantly different from the control, we investigated whether fear or message minimization impacted intentions for those in the self-focus AR × reinforcement condition. First, a mediation model was used to test whether self-focus AR × reinforcement affects perceived fear through perceived threat susceptibility and perceived threat severity (Figure 14). A significant indirect effect of self-focus AR × reinforcement on fear was found with perceived threat severity as the only mediator (b=.16, 95% CI 0.05-0.29, SE 0.06) and with perceived threat susceptibility as the only mediator (b=.13, 95% CI 0.04-0.22, SE 0.05). In addition, a significant indirect effect of self-focus AR × reinforcement on fear was found when both perceived threat susceptibility and threat severity were included as serial mediators (b=.15, 95% CI 0.05-0.28, SE 0.06).

Next, a mediation model was used to test whether the self-focus AR × reinforcement condition affects behavioral intention through fear. A significant indirect effect of the self-focus AR × reinforcement condition on intention was found with fear as the mediator (b=.07, 95% CI 0.01-0.16, SE 0.04). An additional model tested whether the self-focus AR × reinforcement condition affects message minimization through fear, threat severity, or threat severity. A significant negative indirect effect of the self-focus AR × reinforcement condition on message minimization was found with severity as the mediator (b=–.07, 95% CI –0.16 to –0.008, SE 0.04. A negative serial mediation effect with susceptibility and severity was also found (b=–.07, 95% CI –0.16 to –0.008, SE 0.04).

Figure 13. The self-focus augmented reality (AR) × reinforcement, susceptibility, severity, and intention mediation model. The self-focus AR × reinforcement condition resulted in an indirect effect on intention through threat susceptibility and threat severity. This condition also had an indirect effect on perceived threat severity through perceived threat susceptibility.

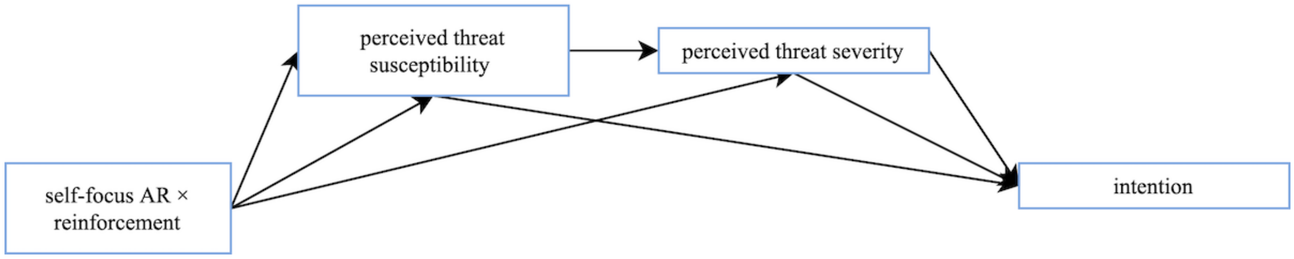
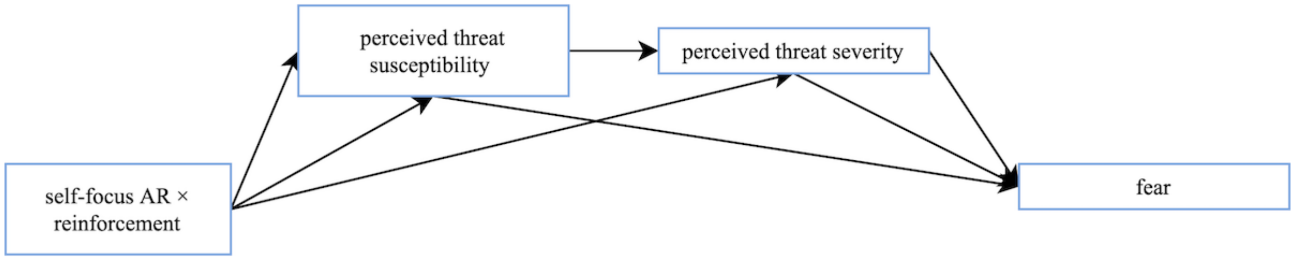


Figure 14. The self-focus augmented reality (AR) × reinforcement, susceptibility, severity, and fear mediation model. The self-focus AR × reinforcement condition resulted in an indirect effect on fear through threat susceptibility and threat severity.



Effects of Using an Avatar on Outcome Expectancy, Fear, Threat Severity, Threat Susceptibility, and Message Minimization (Hypothesis 2)

Hypothesis 2 proposed that combining vicarious reinforcement and self-focused AR while using an avatar would result in higher levels of positive outcome expectancies, fear, perceived threat severity, perceived threat susceptibility, and message minimization when compared to a control. Compared to the control, those in the avatar condition did not have significantly different levels of outcome expectancy (*P*=.42), severity (*P*=.49), susceptibility (*P*=.15), fear (*P*=.23), or message minimization (*P*=.17).

User Feedback

At the end of the questionnaire, we asked participants to enter optional free-form text about the study. Themes among the responses included impact on knowledge, risk perception formation, and challenges with the self-focus AR × reinforcement condition. Several participants in the reinforcement condition provided generalized statements that the study was educational and helpful. The following comments from the self-focused AR group provided more details.

> *I thought the handwashing animation together with the illustration of where dirt is cleaned from the hands was very informative.* [P1]

Some participants learned a new technique or strategy:

> *I learned some new handwashing techniques! (Particularly, locking your hands together by curling your fingers into each other to get the backs of the fingers).* [P2]
>
> *I hadn't thought about separately lathering and washing my thumbs.* [P3]

Comments on personal concerns about risk indicate that future work measuring these variables may consider the time spent around others vs alone and personal risk:

> *I answered questions knowing that my husband's and my job allow us to work from home, which decreases our risk significantly, and that most of my family lives in a rural area, also less susceptible to infection.* [P4]
>
> *I know that I take it way more seriously due to the cancer treatment drug I take than most of my friends and peers because if I get it, I am not strong enough to fight it off. I think that factors in way more than friends and family risk, at least for me personally.* [P5]

Lastly, a few participants in the self-focus AR × reinforcement group expressed confusion about the design. P7, for example, expressed difficulty in positioning themselves on the screen.

> *It was an interesting survey, but the instruction given for the hand part was kind of hard to complete because the outline of the hands and the picture did not match. However, I tried my best to make it work.* [P7]
>
> *Found the movements in the video hard to follow along with, but I tried my best!* [P8]
>
> *The handwashing directions confused me. At first, I didn't understand that I wasn't supposed to mimic the exact instructions.* [P9]

The responses of P8 and P9 suggest that participants may have practiced along with the video animation. Practicing was not a requirement of participation but appeared to be a trend among those in this condition.

Discussion

This study explored the impact of self-focus and vicarious reinforcement design interventions on psychological predictors of behavior change during the COVID-19 pandemic. Our results showed that combining self-focused AR with vicarious reinforcement increases perceived threat severity and threat susceptibility and could potentially impact behavioral intentions.

Behavior Intention and Outcome Expectancy

Our results did not show any direct effects on behavioral intention (Figure 9). Given the severity of COVID-19, we expected these findings to result from a ceiling effect, as similar studies on behavior messaging strategies have found during COVID-19 [37,38]. A surprising finding was the lack of impact of the design interventions on outcome expectancy; however, this also appears to result from a ceiling effect (Figure 10), possibly resulting from the surge of news messaging surrounding hand hygiene's role in mitigating risk during COVID-19. News segments showed the impact of hand hygiene using blacklights [39] to simulate how hand washing reduces viruses' presence, which may have helped combat any disbelief. Due to the ceiling effect, our findings require replication studies post–COVID-19 or studies that include intention and outcome expectancy measures less susceptible to the ceiling effect to investigate the impact of self-focused AR and vicarious reinforcement.

Perceived Threat Severity, Threat Susceptibility, Fear, and Intention

Our findings also revealed interesting relationships between the self-focus AR × reinforcement condition, perceived threat severity, threat susceptibility, fear, and intention. Mediation models showed the self-focus AR × reinforcement condition to positively affect intention and fear through increased perceived threat susceptibility and threat severity. We found self-focus AR × reinforcement to increase perceived threat severity through increased perceived susceptibility. Lastly, we found self-focus AR × reinforcement to indirectly affect intention with fear as the mediator. These results suggest that design strategies that layer a health threat directly on an individual's reflection may increase one's perceived threat susceptibility, threat, severity, fear, and indirectly behavioral intention. While such strategies might help meet behavior change design objectives, it is essential to note the potential consequences of designs that increase fear, especially in the context of a public health emergency.

Based on the EPPM, Li [22] tested a model for protective behaviors during a public health emergency with a study during the Ebola outbreak of 2014. The study found perceived threat to have a significant effect on fear and fear controls. Our study partially supports these findings, indicating an impact of threat severity on fear but not a significant positive effect of fear on message minimization (a fear control mechanism). The danger/fear control responses and the impact of self-focused AR likely varies for each health behavior context, as levels of fear will be different for each health threat. While our study did not show adverse effects, researchers and designers should still use caution if utilizing similar design techniques to effect behavior change. More research is needed on the adverse effects of fear concerning triggering fear control mechanisms through design interventions. In addition, increased fear could have mental health implications. Harper et al [20] found fear of COVID-19 to be a positive predictor of behavior change and fear to be correlated with decreased physical and environmental quality of life. Given our findings, designers must investigate the extent to which a design strategy that involves self-focused AR with a health threat increases fear.

Independent Use of Self-focused Attention and Vicarious Reinforcement

Self-focused AR and vicarious reinforcement embedded as features independently (versus combined) did not show a significant result on any of the tested predictors of behavior change except fear. We provide a few possible explanations for these results. First, regarding mirror self-focus, threat severity, and susceptibility, there may have been too large of a time gap between when participants reviewed the health information provided and when they looked at their self-reflection. In the conditions combining the features, there was content on the screen, reminding participants of the threat at hand. Future work may account for this difference by providing text-based information over one's reflection. Second, regarding vicarious reinforcement and outcome expectancy, Figure 10 indicates the presence of a ceiling effect. This study may need to be replicated for a different health threat or mitigating behavior that wouldn't have as many positive, strong outcome expectancy beliefs.

Limitations and Future Work

It is important to note that this research took place during a long-term global public health emergency with restrictions on lifestyles that can take a while to get adjusted to. Health perceptions related to current circumstances are subject to change throughout the lifecycle of a pandemic. Our findings warrant replication studies that consider changes in severity, government mandates, social perceptions, and the availability and range of tools for risk mitigation (vaccines/medication, personal protective equipment, etc).

The data used in this study are self-reported and susceptible to response biases, specifically social desirability bias. Due to the severity of the pandemic, government mandates, news coverage, and social discussions may have increased the pressure to respond in ways that align with social norms. Future work should aim to use methods to decrease the impact of this limitation.

It is important to note that focusing on individual constructs may create an ineffective design system if the construct only works in combination with other constructs [40]. This should be taken into consideration as self-focused AR is explored in the future, possibly adding measures for other behavior change constructs such as normative beliefs and social facilitation.

While our study indicates that combining self-focused AR with vicarious reinforcement may affect health behavior change by influencing threat severity and susceptibility, we lack a data-driven explanation of why. Future work may benefit from the inclusion of quantitative measures for self-focused attention to compare with severity and susceptibility scores.

Future work may lend itself to developing experimental methods to explore the extent to which self-monitoring, reflective thinking, self-evaluation, and emotion management naturally occur (or do not occur) when using self-focused AR. Conducting these experiments will provide deeper insights into how self-focused AR impacts the psychological mechanism related to behavior change and possibly inspire experiments on how the combination of self-focused attention and other design features could enhance this effect.

According to user feedback, future work should also consider accounting for time spent around others vs alone and personal health when measuring perceived threat severity. Responses also indicate that when measuring perceptions about new messaging, participants should be instructed to respond based on their preferred news source to limit confusion related to the different opinions they hold for individual new sources.

Two comments from the user feedback indicated that participants might have been actually practicing handwashing movements while viewing the animation. During the screenshot verification process, it was noted whether a participant was observed practicing along with the handwashing video. Practicing was not a requirement of participation and was not mentioned in any instructions provided to them. A total of 32 out of 63 participants were observed practicing along with the video in the self-focus AR × reinforcement condition. The self-focused AR and avatar conditions had 7 and 6 individuals observed practicing, respectively. These results cannot be used to make any claims due to the study's technical setup. Those in the self-focus AR × reinforcement group were instructed to have their hands in view of the camera. Those in other conditions may have practiced off-camera. However, as practicing may affect behavior, this is another potential area for future research.

Conclusion

As self-focused AR technologies grow in popularity, it is important to understand how such experiences could impact perceptions, emotions, and behavioral intentions. Previous research [5,12-14,28] has explored self-focused AR to varying degrees revealing a potential impact on health behavior. Our study expands upon this work by combining self-focused AR and vicarious reinforcement. Doing so helped to reveal insights about the impact of each feature on perceptions and emotions as they relate to behavior change.

We found that displaying germs disappearing directly from the user's self-reflection during a handwashing animation will result in higher scores for perceived threat severity and susceptibility when compared to the control or conditions that implemented self-reflection and a display of germs disappearing separately. Increased perceived severity and susceptibility were found to increase behavioral intention. These findings indicate that combining self-focused AR with vicarious reinforcement may be an effective strategy for health communication designers. However, we also voice concern about the possible adverse effects of heightening levels of fear as a design strategy. While our study did not show concerning results, prior research indicates that heightening fear as a health communication strategy can lead to defensive reactions (versus changing behavior) [22,25] and can lower quality of life [20]. We recommend that this be taken into consideration by designers whenever augmenting self-focused attention with a health threat, especially during a public health emergency, as fear may already be at concerningly high levels. Future research should further investigate the role of fear, perceived threat severity, and threat susceptibility when using self-focused AR in health contexts and design strategies for maintaining the well-being of the user while inspiring behavior change.

Acknowledgments

We thank our Wellesley College HCI Lab colleagues, family, and friends for participating in the pilot testing for the study. We thank Angela Qian and Karl Ivan M San Luis for their feedback on the study, as well as Shaun Digan for proofreading. This material is based on work supported by the National Science Foundation (1814628). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect those of the National Science Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Factor loadings.
[DOCX File , 16 KB - jmir_v23i6e26963_app1.docx]

References

1. Number of mobile augmented reality (AR) active users worldwide from 2019 to 2024 (in billions). Statista. 2021 Feb 23. URL: https://www-statista-com.proxy.library.nyu.edu/statistics/1098630/global-mobile-augmented-reality-ar-users/ [accessed 2021-03-27]
2. Billinghurst M, Clark A, Lee G. A Survey of Augmented Reality. Maryland, USA: Now Foundations and Trends; 2015.
3. Snapchat Lens Studio. Snapchat. 2020. URL: https://lensstudio.snapchat.com [accessed 2021-04-30]
4. About Instagram. Instagram. 2020. URL: https://about.instagram.com/about-us [accessed 2021-04-30]
5. Poh M, McDuff D, Picard R. A medical mirror for non-contact health monitoring. In: ACM SIGGRAPH 2011 Emerging Technologies. New York, NY: Association for Computing Machinery; 2011 Presented at: SIGGRAPH '11: Special Interest Group on Computer Graphics and Interactive Techniques Conference; August; Vancouver, BC. [doi: 10.1145/2048259.2048261]
6. Nissan USA. 2014 Apr 15. URL: https://www.nissanusa.com/experience-nissan/news-and-events/smart-rearview-mirror.html [accessed 2020-12-30]
7. memomi. URL: https://memorymirror.com/ [accessed 2020-12-30]
8. Mirror. URL: https://www.mirror.co/ [accessed 2020-12-30]
9. Duval TS, Silvia PJ, Lalwani N. Self-Awareness & Causal Attribution: A Dual Systems Theory. New York, NY: Springer Science & Business Media; 2012.
10. Wicklund R. Objective self-awareness. In: Berkowitz L, editor. Advances in Experimental Social Psychology. Volume 8. New York, NY: Academic Press; 1975:233-275.
11. Flemings M, Kazmi S, Pak R, Shaer O. Crimson Wave: Shedding Light on Menstrual Health. New York, NY: Association for Computing Machinery; 2018 Presented at: TEI '18: Twelfth International Conference on Tangible, Embedded, and Embodied Interaction; March; Stockholm, Sweden p. 343-348. [doi: 10.1145/3173225.3173292]
12. Colantonio S, Coppini G, Germanese D, Giorgi D, Magrini M, Marraccini P, et al. A smart mirror to promote a healthy lifestyle. Biosystems Engineering 2015 Oct;138:33-43. [doi: 10.1016/j.biosystemseng.2015.06.008]
13. Besserer D, Bäurle J, Nikic A, Honold F, Schüssel F, Weber M. Fitmirror: a smart mirror for positive affect in everyday user morning routines. In: Proceedings of the Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction. New York, NY: Association for Computing Machinery; 2016 Presented at: ICMI '16: International Conference on Multimodal Interaction; November; Tokyo, Japan p. 48-55. [doi: 10.1145/3011263.3011265]
14. Jung S, Lee J, Biocca F, Kim JW. Augmented Reality in the Health Domain: Projecting Spatial Augmented Reality Visualizations on a Perceiver's Body for Health Communication Effects. Cyberpsychol Behav Soc Netw 2019 Mar;22(2):142-150. [doi: 10.1089/cyber.2018.0028] [Medline: 30668138]
15. Bandura A. Social Cognitive Theory of Mass Communication. Media Psychology 2001 Aug;3(3):265-299. [doi: 10.1207/s1532785xmep0303_03]
16. Aiello AE, Coulborn RM, Perez V, Larson EL. Effect of Hand Hygiene on Infectious Disease Risk in the Community Setting: A Meta-Analysis. Am J Public Health 2008 Aug;98(8):1372-1381. [doi: 10.2105/ajph.2007.124610]
17. Your Health: If You Are Sick or Caring for Someone. Centers for Disease Control and Prevention. 2020 Feb 11. URL: https://www.cdc.gov/coronavirus/2019-ncov/if-you-are-sick/index.html [accessed 2020-02-01]
18. 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]
19. Ajzen I. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior 1. J Appl Soc Psychol 2002;32(4):665-683. [doi: 10.1111/j.1559-1816.2002.tb00236.x]

20. Harper CA, Satchell LP, Fido D, Latzman RD. Functional Fear Predicts Public Health Compliance in the COVID-19 Pandemic. Int J Ment Health Addict 2020 Apr 27:1-14 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]
21. Brewer NT, Chapman GB, Gibbons FX, Gerrard M, McCaul KD, Weinstein ND. Meta-analysis of the relationship between risk perception and health behavior: the example of vaccination. Health Psychol 2007 Mar;26(2):136-145. [doi: 10.1037/0278-6133.26.2.136] [Medline: 17385964]
22. Li X. Media exposure, perceived efficacy, and protective behaviors in a public health emergency. Int J Commun 2018;12:20 [FREE Full text]
23. Ferrer RA, Klein WM. Risk perceptions and health behavior. Curr Opin Psychol 2015 Oct 01;5:85-89 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.012] [Medline: 26258160]
24. Lerner JS, Keltner D. Beyond valence: Toward a model of emotion-specific influences on judgement and choice. Cognition & Emotion 2000 Jul;14(4):473-493. [doi: 10.1080/026999300402763]
25. Witte K. Putting the fear back into fear appeals: The extended parallel process model. Communication Monographs 1992 Dec;59(4):329-349. [doi: 10.1080/03637759209376276]
26. Hill WF. Learning theory and the acquisition of values. Psychol Rev 1960 Sep;67(5):317-331. [doi: 10.1037/h0043097] [Medline: 13714324]
27. Bandura A, Ross D, Ross SA. Vicarious reinforcement and imitative learning. Journal of Abnormal and Social Psychology 1963;67(6):601-607. [doi: 10.1037/h0045550]
28. Yee N, Bailenson J. The Proteus effect: The effect of transformed self-representation on behavior. Human Comm Res 2007 Jul;33(3):271-290. [doi: 10.1111/j.1468-2958.2007.00299.x]
29. Fox J, Bailenson JN. Virtual Self-Modeling: The Effects of Vicarious Reinforcement and Identification on Exercise Behaviors. Media Psychology 2009 Mar 05;12(1):1-25. [doi: 10.1080/15213260802669474]
30. Infection prevention and control. World Health Organization. URL: https://www.who.int/teams/integrated-health-services/infection-prevention-control [accessed 2021-01-02]
31. Prolific. 2020. URL: https://www.prolific.co [accessed 2021-04-30]
32. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, NY: Guilford Publications; 2017.
33. Python. Python Software Foundation. URL: https://www.python.org/ [accessed 2020-12-30]
34. McKinney W. Data Structures for Statistical Computing in Python. 2010 Presented at: Proceedings of the 9th Python in Science Conference; June 28 - July 3; Austin, Texas p. 56-61 URL: https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf [doi: 10.25080/majora-92bf1922-00a]
35. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Polat, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar 03;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
36. IBM SPSS Statistics for Windows, Version 27. IBM Corp. 2020. URL: https://www.ibm.com/analytics/spss-statistics-software [accessed 2020-12-30]
37. Everett J, Colombatto C, Chituc V, Brady W, Crockett M. The effectiveness of moral messages on public health behavioral intentions during the COVID-19 pandemic. psyarxiv.com 2020:1-23. [doi: 10.31234/osf.io/9yqs8]
38. Barari S, Caria S, Davola A. Evaluating COVID-19 public health messaging in Italy: Self-reported compliance and growing mental health concerns. medRxiv. Preprint posted online April 5, 2020 [FREE Full text] [doi: 10.1101/2020.03.27.20042820]
39. Krauth D. Coronavirus prevention: Are your hands clean? A black light experiment. ABC 7 News. 2020 Mar 15. URL: https://abc7news.com/6004151/ [accessed 2020-12-30]
40. Hekler E, Klasnja P, Froehlich J, Buman M. Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. : Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2013; Paris, France p. 3307-3316. [doi: 10.1145/2470654.2466452]

Abbreviations

AR: augmented reality
CDC: Centers for Disease Control and Prevention
EPPM: Extended Parallel Process Model
WHO: World Health Organization

Edited by C Basch; submitted 05.01.21; peer-reviewed by J Lee, D Bowen; comments to author 23.02.21; revised version received 06.04.21; accepted 16.04.21; published 29.06.21.

Please cite as:
Seals A, Olaosebikan M, Otiono J, Shaer O, Nov O
Effects of Self-focused Augmented Reality on Health Perceptions During the COVID-19 Pandemic: A Web-Based Between-Subject Experiment
J Med Internet Res 2021;23(6):e26963
URL: https://www.jmir.org/2021/6/e26963
doi:10.2196/26963
PMID:33878017

Original Paper

Preferences for Accessing Medical Information in the Digital Age: Health Care Professional Survey

Evelyn R Hermes-DeSantis[1,2*], PharmD; Robert T Hunter[3*], PharmD; Julie Welch[4*], RN, MSN, ANP-BC; Roma Bhavsar[4*], PharmD; Daniel Boulos[5*], MBA, PharmD; Marie-Ange Noue[6*], PhD

[1]phactMI, West Point, PA, United States
[2]Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, Piscataway, NJ, United States
[3]EMD Serono Inc, Rockland, MA, United States
[4]Biogen, Cambridge, MA, United States
[5]Bristol Myers Squibb, Lawrenceville, NJ, United States
[6]EMD Serono, Mississauga, ON, Canada
[*]all authors contributed equally

Corresponding Author:
Evelyn R Hermes-DeSantis, PharmD
phactMI
PO Box 320
West Point, PA, 19486
United States
Phone: 1 215 588 1585
Email: evelyn@phactmi.org

Abstract

Background: Health care professionals (HCPs) routinely have questions concerning the medications they are recommending. There are numerous resources available; however, each has its own advantages and disadvantages.

Objective: The purpose of this survey was to gain knowledge of the preferred methods and sources HCPs use to obtain information concerning medications.

Methods: A total of 511 HCPs (202 physicians, 105 pharmacists, 100 advance practice nurses, 53 registered nurses, and 51 physician assistants) were surveyed through a third-party market research firm. All participants were practicing in the United States. Individuals working for a pharmaceutical company were excluded. The survey collected demographics, frequency of searching medical information, types of questions searched, sources of medical information, and rationale for preferred and nonpreferred sources of medical information. Use of medical information resources were rated on a 5-point ordinal scale. Data were analyzed with descriptive statistics.

Results: Of the 511 respondents, 88.5% (452/511) searched for medical information either daily or several times per week. The most common questions involved dosing and administration, drug-drug interactions, adverse events and safety, clinical practice guidelines, and disease state information. The main rationale for using specific medical websites or apps and general online search engines frequently or very frequently was ease of use (medical websites or apps: 269/356, 75.6%; general online search engines: 248/284, 87.3%). Accuracy was the main rationale for frequent or very frequent use of medical literature search databases (163/245, 66.5%), prescribing labels or information (122/213, 57.3%), and professional literature (120/195, 61.5%). The main reason for rarely or never using specific medical websites or apps and medical literature search databases was unfamiliarity (medical websites or apps: 16/48, 33%; medical literature search databases: 35/78, 45%); for general online search engines, inaccuracy (34/54, 63%); and for prescribing labels or information and professional literature, excessive time (prescribing labels or information : 54/102, 52.9%; professional literature: 66/106, 62.3%). The pharmaceutical company was sometimes used as a resource for medical information. When the medical information department was used, the call center and the website were considered thorough and complete (call center: 14/25, 56%; website: 33/55, 60%). However, the rationale for not using the call center was the time required (199/346, 57.5%) and the website being unfamiliar (129/267, 48.3%).

Conclusions: The driving forces in the selection of resources are accuracy and ease of use. There is an opportunity to increase awareness of all the appropriate resources for HCPs which may aid in their daily clinical decisions. Specifically, pharmaceutical company medical information departments can help fulfill this need by addressing two major challenges with use of the pharmaceutical company: lack of awareness of medical information services and the speed at which responses are disseminated. Overall, there is lack of understanding or appreciation of the range of pathways to obtain published information and knowledge from pharmaceutical company medical information services. Among the many challenges resource champions will face are the ability to effectively make resources and their platforms accessible, known, and useful to the scientific community.

(J Med Internet Res 2021;23(6):e25868) doi:10.2196/25868

KEYWORDS

information-seeking behavior; access to information; internet; physicians; nurses; pharmacists; medical literature; databases; search tools; medical information

Introduction

Health care professionals (HCPs) routinely seek medical information concerning the therapies recommended or used to manage and treat their patients. The US Food and Drug Administration (FDA) approved over 100 novel medications between 2019 and 2020 with a trend for increasing annual approvals over the last decade [1]. Meanwhile, medical information resources that HCPs use to address questions and issues in caring for their patients are growing and expanding [2].

HCPs have a wide variety of options for seeking answers to their questions. The number of resources has expanded and now includes drug aggregate platforms (Epocrates, MicroMedex, Up-to-Date, Medscape), medical information departments, professional journals, prescribing labels, electronic health record systems, textbooks, search engine websites, and academic drug information centers [2]. Although the drug aggregate platforms provide a wealth of information, these databases may contain misinformation [3]. In a 2020 study of neurologists, online resources were preferred (96%) compared to offline resources (47%) [4]. According to the 2013 study by Kritz [5], physicians used online resources for medical information; however, access to quality information was a barrier of note.

Medical information services offered by pharmaceutical companies can provide evidence-based, scientifically balanced, accurate, truthful, nonmisleading responses to unsolicited inquiries from HCPs. These unsolicited inquires can be questions concerning the FDA-approved product labels or questions beyond the labeled information. The responses to these inquiries conform with internal procedures and policies as well as with the FDA draft guidance document [6].

Previous literature has suggested that ease of access and quality are important factors when physicians or medical students search for medical information [7]. For example, in a 2009 survey, 92% of physicians reported clicking results toward the top of a page when searching for medical information online [8]. Search engine algorithms are continually updating and adapting in an effort to improve the search results, identify high quality content, and devalue lower-quality content [9]. However, online access alone may not be enough, as barriers such as time constraints confounded by password or account creation requirements on some medical information department websites may impede use of these resources. Another key component noted in past surveys is that HCPs consider current information from a trusted source as high-quality information [7]. Overall awareness, access, and trust have been discussed as important factors HCPs consider when deciding on medical information resources [2]. In a comprehensive literature review of more than 30 studies, Davies et al [10] found that the majority of research focused predominantly on physicians and the type of information they sought. A key barrier to physicians' search for information was the time needed to perform the search effectively. The studies on this topic [2,7-10], most of which are outdated, highlight the lack of published literature on this subject, particularly evaluations on the search preferences of HCPs apart from those of physicians.

Given the numerous resources HCPs can use today for searching or requesting medical information, there is a need to better understand their search preferences, processes, and barriers to using these resources. Therefore, the purpose of this study was to gain knowledge about the frequency, preferred methods, and most common sources HCPs use to obtain medical information. An additional goal was to evaluate and categorize the rationale of HCP choices with the aim of enhancing medical information services within pharmaceutical companies.

Methods

A deidentified, web-based, qualitative survey was designed by members of phactMI to collect the opinions of HCPs concerning their search preferences for medical information. Surveys were distributed through a third party platform (Dynata market research organization). Recruitment and participation were communicated via email through Dynata. Participants received a unique identifier that did not reveal their identity to the study team. The survey was administered once and was open for 1 week in March 2019.

HCPs (physicians, pharmacists, nurse practitioners or advance practice nurses, registered nurses, and physician assistants) received an email inviting them to participate in the survey. These HCPs were registered with the third-party surveying platform, Dynata. The survey was distributed to a panel of verified HCPs and was available until the prespecified convenience sample was obtained. A sample size of 202 physicians, 105 pharmacists, 100 advance practice nurses, 53 registered nurses, and 51 physician assistants was chosen and felt to be representative of those HCPs who contact the typical mid- and large-size pharmaceutical companies' medical information departments. Registered HCPs responded to qualifying questions based on the inclusion and exclusion criteria. Participants were included based on their profession (practicing physician, pharmacist, nurse practitioner or advanced practice nurse, registered nurse, and physician assistant) and their country of practice (the United States). Participants who were not currently practicing or who were employed by a pharmaceutical or biopharmaceutical company were excluded. Once 511 qualified respondents completed the survey, the recruitment period ended. Participants were asked for their consent to participate in the survey prior to survey administration. Dynata follows the International Chamber of Commerce/European Society for Opinion and Marketing Research (ICC/ESOMAR) International Code on Market, Opinion, and Social Research and Data Analytics.

The survey collected information regarding HCPs' demographics (HCP type, years of practice, specialty, and practice setting) and their search preferences and processes. The survey took about 10 minutes and consisted of 13 questions. Questions in the survey included frequency of searching for medical information, inquiry categories, sources of medical information used, and rationale for preferred and nonpreferred sources of medical information. The survey consisted of multiple choice questions, ranking, and, to a lesser extent, a free-text field. Based on responses given throughout the survey, additional information was ascertained regarding reasons why or why not certain sources of medical information were preferred or not preferred. See Multimedia Appendix 1 for the full list of survey questions and available responses.

Descriptive analysis of data was conducted. Data were analyzed based on the entire cohort as well as on profession, specialty, and years in practice. Chi-square and Fisher exact tests were used for categorical data.

Results

Demographics

Based on the convenience sample set, a total of 511 health care professionals in the United States were included in the survey, comprising physicians (202/511, 39.5%), pharmacists (105/511, 20.5%), nurse practitioners or advanced practice nurses (100/511, 19.6%), registered nurses (53/511, 10.4%), and physician assistants (51/511, 10.0%). The most common practice settings represented included private practice (211/511, 41.3%), community hospitals (127/511, 24.9%), and academic or teaching hospitals (102/511, 20.0%). Consistent with the dominant practice settings in this survey, almost half (249/511, 48.7%) of all respondents worked in primary care. Moreover, 80.2% (410/511) of those surveyed had been in practice for ≥ 11 years. See Table 1 for additional details.

Table 1. Demographics.

Practice setting[a]	All HCPs[b], n (%)	Physicians, n (%)	Pharmacists, n (%)	NP/APN[c], n (%)	RN[d], n (%)	PA[e], n (%)
All settings	511 (100)	202 (39)	105 (21)	100 (20)	53 (10)	51 (10)
Private practice	211 (41)	132 (65)	1 (1)	41 (41)	8 (15)	29 (57)
Academic/teaching hospital	102 (20)	40 (20)	16 (15)	24 (24)	13 (25)	9 (18)
Community hospital	127 (25)	60 (30)	21 (20)	21 (21)	17 (32)	8 (16)
Other[f]	204 (40)	33 (16)	84 (80)	26 (26)	19 (36)	14 (27)
Specialty practice						
Primary care	249 (49)	102 (50)	50 (48)	57 (57)	12 (23)	28 (55)
Oncology/hematology	36 (7)	10 (5)	13 (12)	4 (4)	7 (13)	2 (4)
Cardiology	25 (5)	11 (5)	0 (0)	7 (7)	5 (9)	2 (4)
Psychiatry	17 (3)	11 (5)	2 (2)	3 (3)	1 (2)	0 (0)
Orthopedics	10 (2)	5 (2)	0 (0)	2 (2)	0 (0)	3 (6)
General surgery	10 (2)	6 (3)	1 (1)	0 (0)	3 (6)	0 (0)
Endocrinology	9 (2)	3 (1)	1 (1)	3 (3)	1 (2)	1 (2)
Pulmonology	5 (1)	5 (2)	0 (0)	0 (0)	0 (0)	0 (0)
Neurology	5 (1)	4 (2)	0 (0)	1 (1)	0 (0)	0 (0)
Rheumatology	4 (1)	7 (3)	1 (1)	0 (0)	0 (0)	1 (2)
Other	141 (28)	43 (21)	37 (35)	23 (23)	24 (45)	14 (27)
Years in practice						
HCPs >20 years	230 (45)	90 (45)	62 (59)	35 (35)	30 (57)	13 (25)
HCPs 11-20 years	180 (35)	68 (34)	32 (30)	40 (40)	14 (26)	26 (51)
HCPs <11 years	101 (20)	44 (22)	11 (10)	22 (22)	8 (15)	12 (23)

[a] More than one practice setting could be selected.

[b] HCPs: health care professionals.

[c] NP/APN: nurse practitioner/advanced practice nurse.

[d] RN: registered nurse.

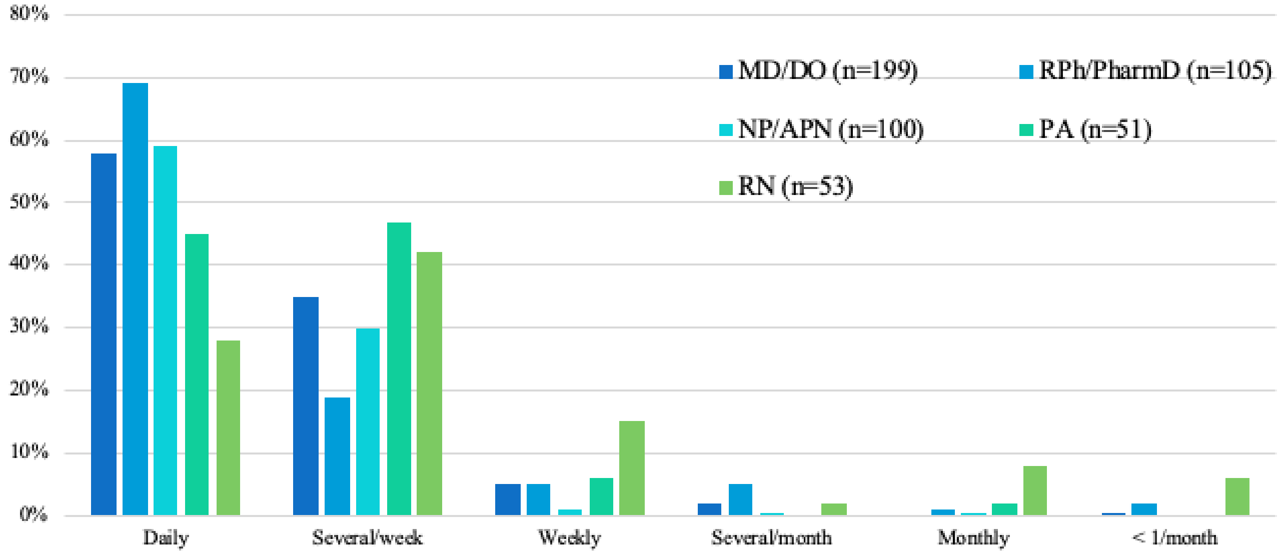
[e] PA: physician assistant.

[f] Other practice settings included health maintenance organization (n=73), pharmacy hospital (n=23), pharmacy retail (n=64), managed care (n=6), research (n=4), long-term care (n=22), and other (n=12).

Frequency of Search

Of the 511 respondents, 452 (88.5%) searched for medical information either daily or several times per week (Figure 1). In particular, 92.6% (187/202) of physicians surveyed search for medical information either daily or several times per week, compared to 87.6% (92/105) of pharmacists, 89.0% (89/100) of advance practice nurses, 70% (37/53) of registered nurses, and 92% (47/51) of physician assistants (X^2_4=49.51; P<.001). Overall, 90.1% (91/101) of respondents with 10 years of practice or less and 88.0% (361/410) of respondents with 11 or more years of practice searched for medical information daily or several times per week.

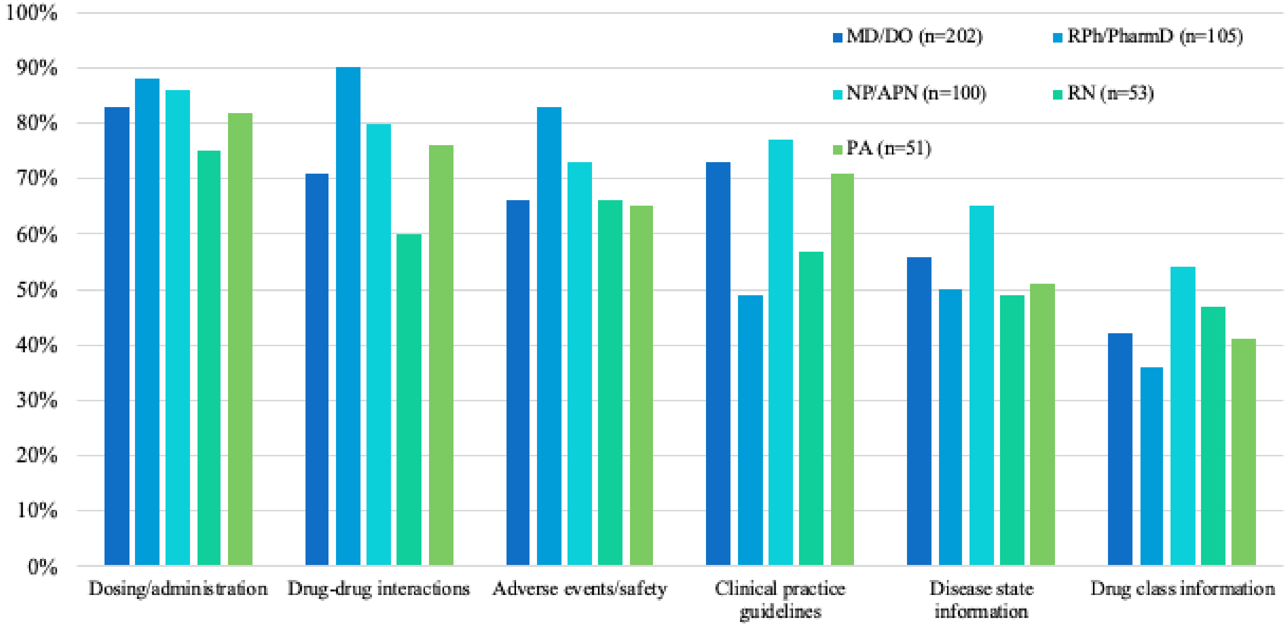
Figure 1. Frequency of medical information searches. APN: advanced practice nurse; DO: Doctor of Osteopathic Medicine; MD: medical doctor; NP: nurse practitioner; PA: physician assistant; PharmD: Doctor of Pharmacy; RN: registered nurse; RPh: registered pharmacist.



Inquiry Type

The most common questions across all HCP types concerned dosing or administration (428/511, 83.8%), drug-drug interactions (389/511, 76.1%), adverse events and safety (361/511, 70.6%), clinical practice guidelines (342/511, 66.9%), and disease state information (283/511, 55.4%; Figure 2). Overall, there were no significant differences between the professions and the type of information they were looking for.

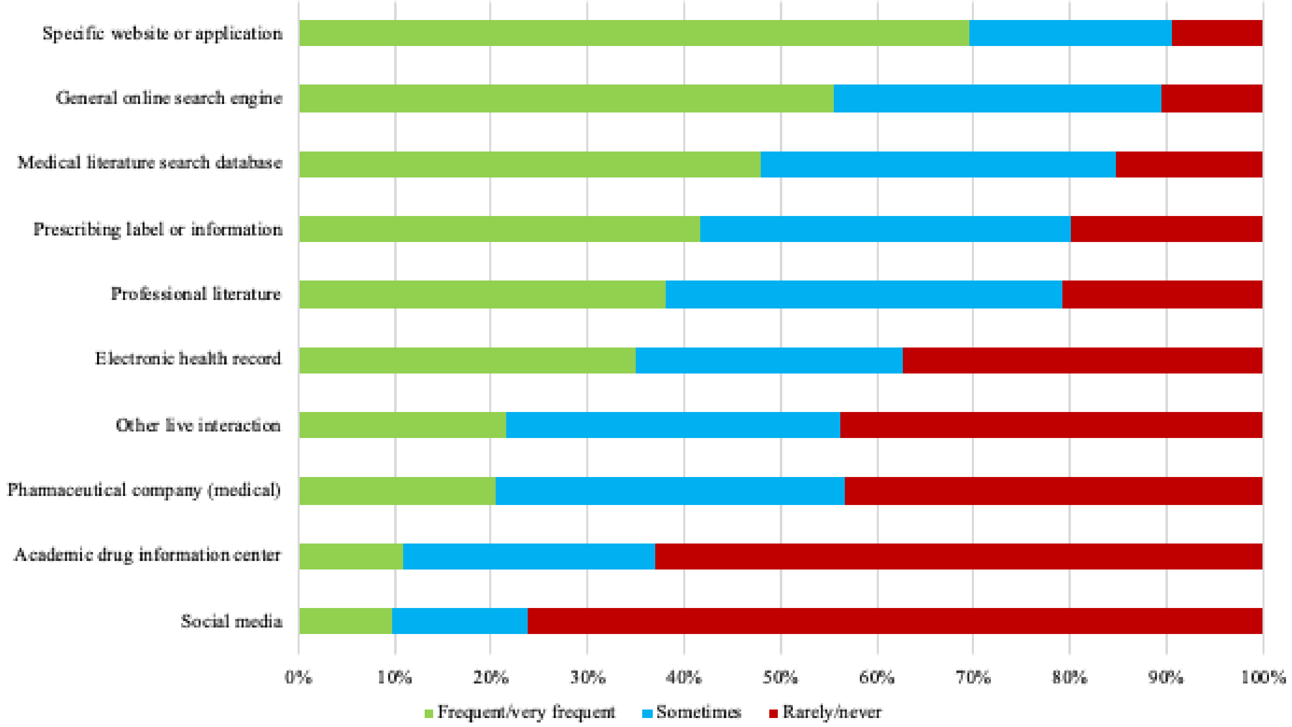
Figure 2. Information typically sought by healthcare professionals. APN: advanced practice nurse; DO: Doctor of Osteopathic Medicine; MD: medical doctor; NP: nurse practitioner; PA: physician assistant; PharmD: Doctor of Pharmacy; RN: registered nurse; RPh: registered pharmacist.



Search Tools

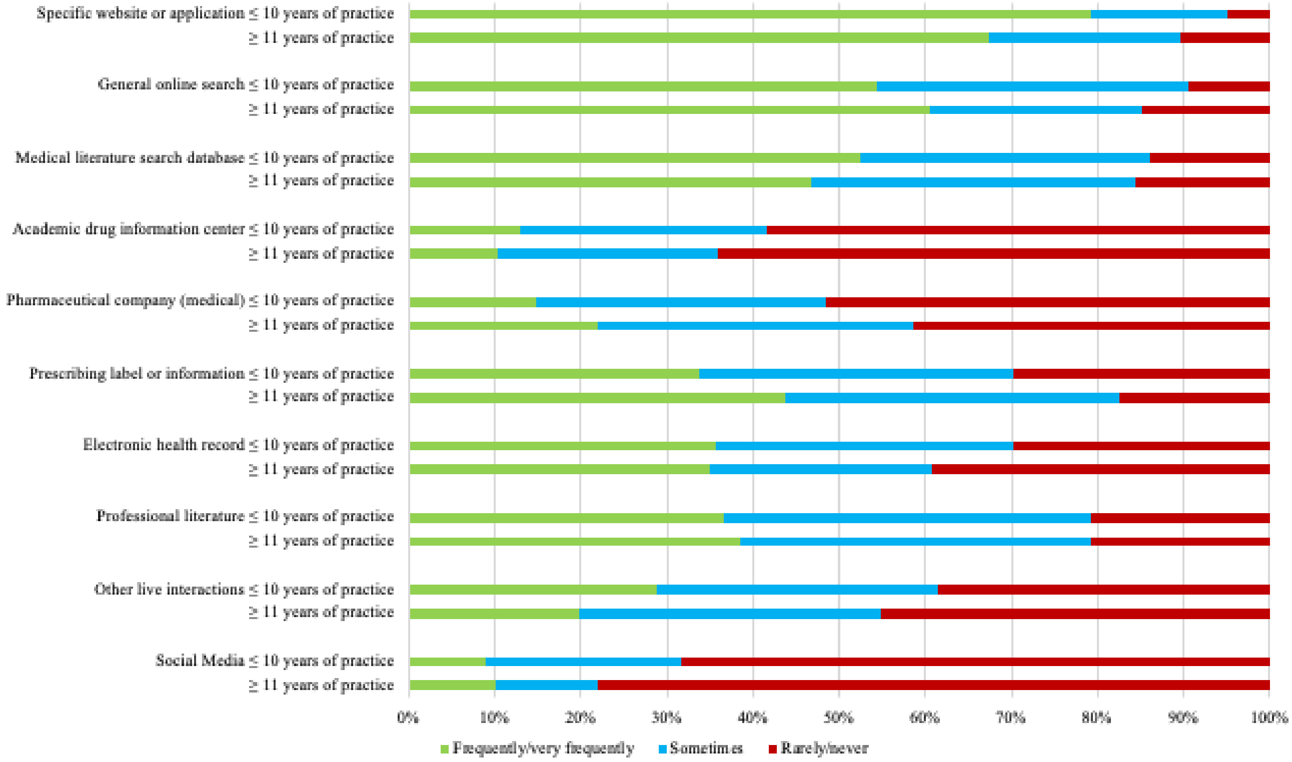
More than half (58.7%, 300/511) of respondents indicated searching for medical information from a desktop, laptop, or workstation a majority (ie, >50%) of the time as opposed to the 25.6% (131/511) of respondents who used a mobile device the majority of the time. Specific medication websites or apps were frequently or very frequently searched by 69.7% (356/511) of the respondents. Furthermore, 55.6% (284/511) of HCPs surveyed use general online search engines such as Google or Yahoo frequently or very frequently. Medical literature search databases, prescribing labels or information, professional literature, and company resources were accessed less frequently (Figure 3).

Figure 3. Frequency of use of major information sources.



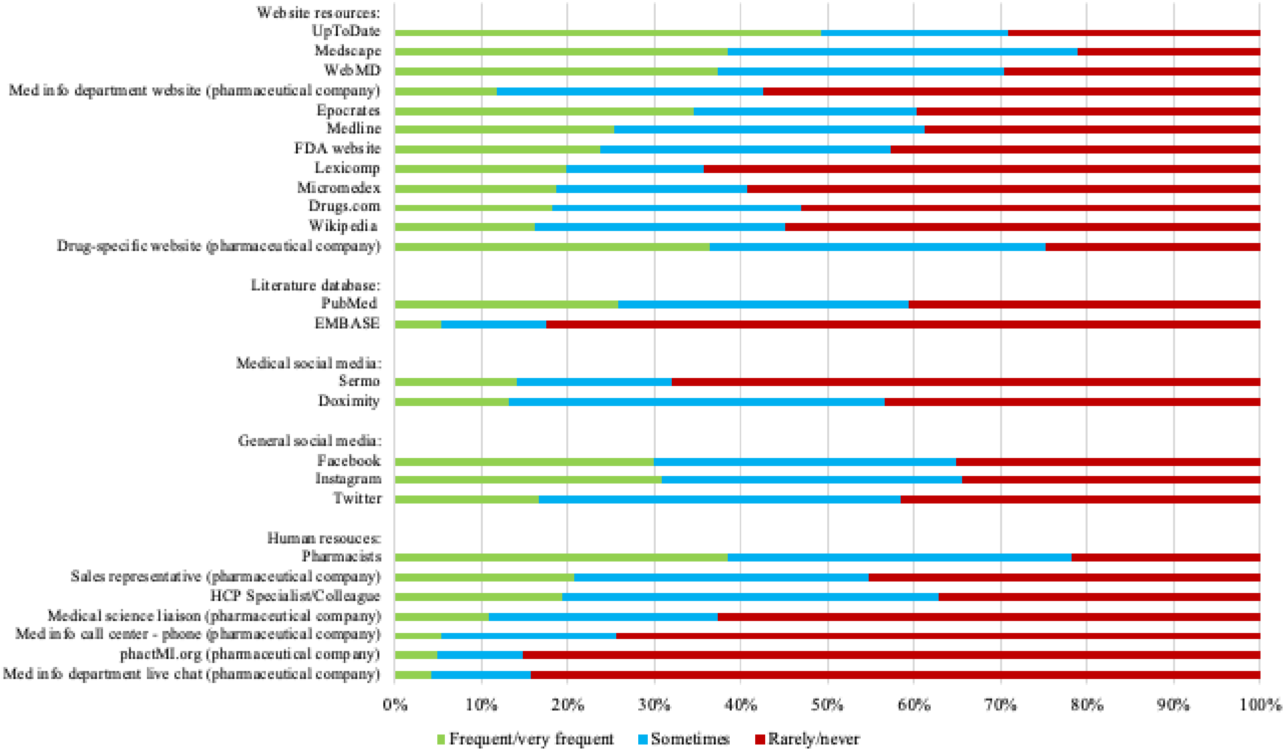
Differences in the search option used were observed based on the number of years in practice. Overall, 24.8% (25/101) of the less experienced respondents (10 years of practice or less) versus 36.8% (151/410) of the respondents with 11 years or more of practice reported sometimes using general online search engines; 79.2% (80/101) of the less experienced respondents versus 67.3% (276/410) of the more experienced respondents reported frequent or very frequent use of specific websites or app for their searching needs; 29.7% (30/101) of the less experienced respondents versus 39.3% (161/410) of the more experienced respondents reported rarely or never using electronic health record information; and 67.3% percent (69/101) of the less experienced respondents reported rarely or never using social media versus 78.0% (320/410) of the more experienced group (Figure 4). Differences in the search option used stratified by the number of years in practice did not reach statistical significance.

Figure 4. Search option preference based on years of practice.



In the analysis regarding the frequency of use of search options, UpToDate ranked highest among all individual sources of medical information (excluding general online search engines). The drug-specific website was the most frequently used resource by those contacting the pharmaceutical company directly. PubMed and MEDLINE were the most popular literature search databases used. Facebook, Sermo, and Instagram were the most popular social media platforms used (Figure 5).

Figure 5. Frequency of use of source information. FDA: Food and Drug Administration; HCP: health care professional.



Reasons for Frequent or Infrequent Use of Search Options

Ease of use was the rationale for the frequent or very frequent use of specific medication websites or apps (269/356, 75.6%) and general online search engines (248/284, 87.3%; Table 2). Accuracy was the major rationale for frequent or very frequent use of medical literature search databases (163/245, 66.5%), prescribing labels or information (122/213, 57.2%), and professional literature (120/195, 61.5%). The main reason for infrequent (rarely or never) use of specific medication websites or apps (16/48, 33%) and medical literature search databases (35/78, 44.9%) was unfamiliarity; for general online search engines, inaccuracy (34/54, 63%); and for prescribing labels or information (54/102, 52.9%) and professional literature, time (66/106, 62.2%; Table 2).

In general, there was low usage of the pharmaceutical company (medical information website, medical science liaison, or sales representative). HCPs in the setting of a hospital pharmacy (6/23, 26%) and retail pharmacy (16/64, 25%) were more likely than those in an academic hospital (16/102, 15.6%) to use the services of the pharmaceutical company frequently or very frequently. When the medical information department was used, the medical information department call center and the website were considered thorough or complete (14/25, 56% and 33/55, 60%); however, the barriers included excessive time (199/346, 57.5%; 79/267, 29.6%) or unfamiliarity with the service (118/346, 34.1%; 129/267, 48.3%; Table 2). Similar findings were reported regarding the pharmaceutical company field medical teams. For HCPs who frequently or very frequently contact a sales representative when searching for medical information, accessibility of the sales representative was the most stated rationale provided.

There was generally low usage of the medical information department live chat service, with only 3.7% (19/511) of HCPs reporting that they use this service frequently or very frequently. When used, accessibility was the main driver (13/19, 68%). For the majority of respondents who indicated having rarely or never contacted the pharmaceutical company via live chat, 33.8% (133/393) were unfamiliar with the service and 51.9% (204/393) noted it took too long.

Free-text entries were reviewed. Additional information regarding practice settings have been included under demographics (see Table 1); however, data obtained from other text fields did not provide any useful information and have not been summarized due to the limited range and number of entries.

Table 2. Rationale for frequent or very frequent or rarely/never use (N=511).

Rationale by use frequency	General online search, n (%)	Medical literature database, n (%)	Specific web app, n (%)	Med info[a] call center, n (%)	Med info[a] website, n (%)	Medical science liaison, n (%)	Drug website, n (%)
Frequent/very frequent[b]	284 (56)	245 (48)	356 (70)	25 (5)	55 (11)	50 (10)	169 (33)
No other option	9 (3)	3 (1)	6 (2)	0 (0)	3 (5)	1 (2)	8 (5)
Familiarity	148 (52)	104 (42)	210 (59)	7 (28)	12 (22)	12 (24)	43 (25)
Responsive/quick	177 (62)	57 (23)	154 (43)	11 (44)	16 (29)	19 (38)	58 (34)
Ease of use	248 (87)	100 (41)	269 (76)	9 (36)	24 (44)	19 (38)	92 (54)
Accuracy	53 (19)	163 (67)	226 (63)	11 (44)	27 (49)	24 (48)	70 (41)
Thorough/complete	58 (20)	137 (56)	197 (55)	14 (56)	33 (60)	27 (54)	67 (40)
Accessibility	168 (59)	97 (40)	211 (59)	13 (52)	25 (45)	27 (54)	94 (56)
Rarely/never[b]	54 (11)	78 (15)	48 (9)	346 (68)	267 (52)	290 (57)	115 (23)
Inaccurate	34 (63)	1 (1)	5 (10)	7 (2)	3 (1)	7 (2)	9 (8)
Not available at my organization	5 (9)	18 (23)	8 (17)	16 (5)	30 (11)	56 (19)	8 (7)
Not thorough enough	28 (52)	3 (4)	6 (13)	23 (7)	19 (7)	18 (6)	34 (30)
Difficult to use	3 (6)	21 (27)	3 (6)	57 (16)	37 (14)	26 (9)	12 (10)
Difficult to access	3 (6)	22 (28)	12 (25)	89 (26)	60 (22)	67 (23)	18 (16)
Unfamiliar with method	3 (6)	35 (45)	16 (33)	118 (34)	129 (48)	120 (41)	30 (26)
Takes too long	15 (28)	32 (41)	10 (21)	199 (58)	79 (30)	90 (31)	31 (27)

[a] Med info: medical information.

[b] Percentages in this row are derived from the total number of responses (N=511).

Discussion

Principal Findings

With the lack of recently published information on search preferences for HCPs, this study provides insights into frequency, preferred methods, and commonly used sources that a broad range of HCPs used to obtain medical information. Overall, 88.5% (452/511) of all HCPs, including almost 95% of physicians, search for medical information either daily or several times per week. Data from this study show that HCPs are using only a few of the aforementioned resources, and many valuable assets are being underused. Medical literature search databases, prescribing labels or information, professional literature, and pharmaceutical company resources were used less frequently compared to other resources, such as general online search engines or specific websites and apps. It is important to note here that general search engines may be used by HCPs to access specific web-based resources such as Medscape and WebMD. Potential reasons for low usage can vary from access issues with literature databases, inability to find a specific answer to a question from the label, or a lack of awareness of pharmaceutical resources such as medical information services or access to company representatives. With the rapidly expanding bank of scientific data, it is becoming increasingly important to have access to information from credible sources.

Taking a closer look at the use of the pharmaceutical company as a resource, less frequent use was due in part to perceived barriers which include the length of time it takes to produce data in response to a specific unsolicited question and the unfamiliarity of the HCPs with the medical information service provided. Previous research has documented that HCPs who use medical information services identify the information as trustworthy [11,12]. However, among the HCPs who had not used medical information services, the top 3 reasons were bias (55%), lack of awareness (41%), and lack of transparency (30%) [12]. The technological advancement and rise in new tools in the pharmaceutical and biotech industry have the potential to better address HCPs' needs, provided that some work is done to resolve the existing barriers.

Limitations

One limitation to this study is the inclusion of only US HCPs. In addition, the demographics of the various HCPs, including physicians (202/511, 39.5%), pharmacists (105/511, 20.5%), nurse practitioners or advanced practice nurses (100/511, 19.6%), registered nurses (53/511, 10.4%), and physician assistants (51/511, 10.0%), was not equally distributed and might have skewed the data towards physician preferences. Another limitation is that the survey was qualitative in nature; therefore, some of the terminology used within the questions could have been subject to personal interpretation. For example, the definition of "ease of use" might have been influenced by an individual's subjective perception. Another point to consider is that no information on the current level of digital tool usage was gathered from the respondents, so it is not clear if the results might have been influenced by a low level of familiarity with digital tools in general. Additional insights from a larger-scale study with HCPs worldwide would be of interest for comparison.

It is also important to realize the survey was conducted in March 2019, prior to the current global COVID-19 pandemic. Current remote workflow and reliance on technology could alter the responses of HCPs if the survey were repeated today, and it is unclear what the long-term consequences and duration of these changes will be.

Conclusions

Research has shown that accuracy and ease of use are the driving forces for HCPs in choosing resources for daily use. Data from this study show an opportunity to increase awareness of all the appropriate resources tailored to HCPs which may aid in their daily clinical decisions. In addition, there is an opportunity for medical information departments from pharmaceutical and biotech companies to help fulfill this need by addressing two major challenges in the use of pharmaceutical companies as a resource: lack of awareness of these medical information services and the speed at which responses are disseminated to HCPs. There is a potential opportunity for medical information departments and field medical teams (ie, medical science liaisons) to work together to overcome these perceived barriers through bringing awareness to the service and increasing accessibility to HCPs by emphasizing the point that medical teams are to only provide tailored scientific data in a nonpromotional manner that most suits their work style and demand [6].

With a significant push towards evidence-based medicine, there remains a need for a unified source for medical information that meets the shifting workstyle needs of practitioners. Optimal treatment decisions are incumbent on current, high-quality, nonpromotional data. The results of this research identified gaps in understanding the numerous ways HCPs obtain published data and the limited knowledge of medical information services provided by pharmaceutical and biotech companies. Among the many challenges resource champions will face are the ability to effectively make resources and their platforms accessible, known, and useful for the scientific community. This will further shape and impact the future of patient outcomes.

Conflicts of Interest

ERHD is employed by phactMI (a nonprofit collaboration of medical information leaders from the pharmaceutical industry). RTH and MAN are employed by EMD Serono, and RTH owns EMD Serono stock. DB is employed by Bristol Myers Squibb. JW and RB are employed by and own Biogen stock. RTH, DB, RB, JW, and MAN are members of phactMI.

Multimedia Appendix 1
Survey questions and answer options.
[DOC File , 45 KB - jmir_v23i6e25868_app1.doc]

References

1. Advancing health through innovation: new drug approvals 2020. US Food and Drug Administration. 2021. URL: https://www.fda.gov/media/144982/download [accessed 2021-02-16]
2. Albano D, Soloff A, Heim K, Mavila S. A theory on the relativity of factors impacting the utilization of medical information services from the pharmaceutical industry. Ther Innov Regul Sci 2016 Sep;50(5):554-559. [doi: 10.1177/2168479016640019] [Medline: 30231753]
3. Randhawa AS, Babalola O, Henney Z, Miller M, Nelson T, Oza M, et al. A collaborative assessment among 11 pharmaceutical companies of misinformation in commonly used online drug information compendia. Ann Pharmacother 2016 May;50(5):352-359 [FREE Full text] [doi: 10.1177/1060028016635196] [Medline: 26917822]
4. Demergazzi S, Pastore L, Bassani G, Arosio M, Lonati C. Information needs and information seeking behavior of Italian neurologists: exploratory mixed methods study. J Med Internet Res 2020 Apr 08;22(4):e14979 [FREE Full text] [doi: 10.2196/14979] [Medline: 32181742]
5. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. J Med Internet Res 2013 Jun 26;15(6):e122 [FREE Full text] [doi: 10.2196/jmir.2436] [Medline: 23803299]
6. Guidance for industry: responding to unsolicited requests for off-label information about prescription drugs and medical devices. US Food and Drug Administration. 2011. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/responding-unsolicited-requests-label-information-about-prescription-drugs-and-medical-devices?source=govdelivery [accessed 2020-06-03]
7. Brennan N, Edwards S, Kelly N, Miller A, Harrower L, Mattick K. Qualified doctor and medical students' use of resources for accessing information: what is used and why? Health Info Libr J 2014 Sep;31(3):204-214 [FREE Full text] [doi: 10.1111/hir.12072] [Medline: 25041386]
8. Google. Connecting with physicians online: searching for answers. 2009. URL: www.gstatic.com/ads/research/en/2009_ConnectingwithPhysiciansOnline.pdf [accessed 2020-06-03]

9. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J 2007 Jun;24(2):78-94 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00713.x] [Medline: 17584211]
10. Strzelecki A. Google Medical update: why is the search engine decreasing visibility of health and medical information websites? Int J Environ Res Public Health 2020 Feb 12;17(4):1-13 [FREE Full text] [doi: 10.3390/ijerph17041160] [Medline: 32059576]
11. Young J, Bhavsar R, Byun J, Toscani M, Reilly P, Hermes-DeSantis E. Opinions and trends of health care providers concerning scientific/standard response documents. Ther Innov Regul Sci 2020 Nov;54(6):1388-1397. [doi: 10.1007/s43441-020-00165-9] [Medline: 32458379]
12. Albano D, Follman S, Patel Y, Gantotti S, Reilly P, Hermes-DeSantis E. Enhancing patient care: value of pharmaceutical medical information service. The Journal for Nurse Practitioners 2021 Mar:1-6 [FREE Full text] [doi: 10.1016/j.nurpra.2021.01.009]

Abbreviations

HCP: health care professional
ICC/ESOMAR: the International Chamber of Commerce/European Society for Opinion and Marketing Research

Edited by R Kukafka; submitted 20.11.20; peer-reviewed by S Demergazzi, B Protus; comments to author 31.12.20; revised version received 24.02.21; accepted 04.05.21; published 19.06.21.

Please cite as:
Hermes-DeSantis ER, Hunter RT, Welch J, Bhavsar R, Boulos D, Noue MA
Preferences for Accessing Medical Information in the Digital Age: Health Care Professional Survey
J Med Internet Res 2021;23(6):e25868
URL: https://www.jmir.org/2021/6/e25868/
doi:10.2196/25868
PMID:36260374

Original Paper

Effect of a Consumer-Focused Website for Low Back Pain on Health Literacy, Treatment Choices, and Clinical Outcomes: Randomized Controlled Trial

Paul William Hodges[1], MedDr, PhD, DSc; Leanne Hall[1], PhD; Jenny Setchell[1], PhD; Simon French[2], PhD; Jessica Kasza[3], PhD; Kim Bennell[4], PhD; David Hunter[5], PhD; Bill Vicenzino[1], PhD; Samuel Crofts[6], MPH; Chris Dickson[7], BA(Hons); Manuela Ferreira[5], PhD

[1]School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia
[2]Department of Chiropractic, Macquarie University, Sydney, Australia
[3]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[4]Centre for Health Exercise and Sports Medicine, Department of Physiotherapy, University of Melbourne, Melbourne, Australia
[5]Institute of Bone and Joint Research, The Kolling Institute, Northern Clinical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia
[6]Centre for Epidemiology and Biostatistics, The University of Melbourne, Melbourne, Australia
[7]Department of Integrative Medicine, Chris O'Brien Lifehouse Hospital, Sydney, Australia

Corresponding Author:
Paul William Hodges, MedDr, PhD, DSc
School of Health and Rehabilitation Sciences
The University of Queensland
Bldg 87A
Brisbane, 4072
Australia
Phone: 61 404 854 589
Fax: 61 733652008
Email: p.hodges@uq.edu.au

Abstract

Background: The internet is used for information related to health conditions, including low back pain (LBP), but most LBP websites provide inaccurate information. Few studies have investigated the effectiveness of internet resources in changing health literacy or treatment choices.

Objective: This study aims to evaluate the effectiveness of the MyBackPain website compared with unguided internet use on health literacy, choice of treatments, and clinical outcomes in people with LBP.

Methods: This was a pragmatic, web-based, participant- and assessor-blinded randomized trial of individuals with LBP stratified by duration. Participants were randomly allocated to have access to the evidence-based MyBackPain website, which was designed with input from consumers and expert consensus or unguided internet use. The coprimary outcomes were two dimensions of the Health Literacy Questionnaire (dimension 2: "having sufficient information to manage my health;" dimension 3: "actively managing my health;" converted to scores 1-100) at 3 months. Secondary outcomes included additional Health Literacy Questionnaire dimensions, quality of treatment choices, and clinical outcomes.

Results: A total of 453 participants were recruited, and 321 (70.9%) completed the primary outcomes. Access to MyBackPain was not superior to unguided internet use on primary outcomes (dimension 2: mean difference −0.87 units, 95% CI −3.56 to 1.82; dimension 3: mean difference −0.41 units, 95% CI −2.78 to 1.96). Between-group differences in other secondary outcomes had inconsistent directions and were unlikely to be clinically important, although a small improvement of unclear importance in the quality of stated treatment choices at 1 month was found (mean difference 0.93 units, 95% CI 0.03 to 1.84).

Conclusions: MyBackPain was not superior to unguided internet use for health literacy, but data suggest some short-term improvement in treatment choices. Future research should investigate if greater interactivity and engagement with the website may enhance its impact.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12617001292369; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=372926

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-027516

(J Med Internet Res 2021;23(6):e27860) doi:10.2196/27860

KEYWORDS

low back pain; randomized controlled trial; internet resources; health literacy

Introduction

Background

Low back pain (LBP) is the leading cause of disability globally [1]. Unnecessary and ineffective management inflates its burden [2], and its impact is worsened by negative messages and beliefs [3,4]. Access to an evidence-based, consumer-focused tool to enhance health literacy and empower active participation in self-management and treatment selection could reduce the burden. Although tools are available, the evaluation of their effectiveness is limited.

The internet is a primary source of health-related information [5-7]. Up to 68% of individuals search for health-related information on the web [7,8]. Information is sought on the internet for treatment decision support [7,9,10], self-management advice [7,9], guidance regarding health care providers [10], increased knowledge [10,11], and preparation for consultations [10]. The advantages of the internet are accessibility, high reach, low cost, and scalability [12]. Unfortunately, most LBP websites are rated poorly [13-15], do not meet consumers' needs [16], provide inaccurate information and treatment recommendations [17], and use inappropriate language [14]. Although systematic reviews provide modest evidence of the efficacy of internet-delivered *interventions* (eg, cognitive behavioral therapy [18,19]), whether *information* resources improve outcomes and behaviors is unknown.

Objectives

This study aims to evaluate access to an LBP website [20] that integrates evidence-based LBP information developed through consumer consultation and expert consensus [21]. We hypothesize that the MyBackPain website more effectively improves health literacy (primary outcome), choice of evidence-based treatments, and clinical outcomes in people with LBP than unguided use of internet resources.

Methods

Trial Design

We conducted a two-arm pragmatic, web-based, superiority randomized controlled trial. It was prospectively registered (Australian New Zealand Clinical Trials Registry ACTRN12617001292369), and the protocol has been published [22]. Ethical approval was obtained from the University of Queensland Ethics Committee (#2017000995). Written informed consent was obtained from all participants.

Participants

We recruited community participants with current LBP using newsletters, email lists, consumer groups, websites, social media, and newsletters to members of a health insurer (Medibank Private). Inclusion criteria were current LBP of any duration, aged 18 years and above (no upper age limit), current residence in Australia, adequate English to use the MyBackPain website, and internet access. Participants were excluded if they reported a previous or existing serious spinal pathology (fracture, cancer, or infection) or specific diagnosis including sciatica (participants with leg pain but no diagnosis of sciatica were not excluded), lumbar spinal stenosis, or nerve root compromise.

Procedure

This trial was conducted on the web. Potential participants were provided with a web link to the participant information sheet and consent form. Those who consented completed an eligibility screening form and, if eligible, were invited to complete the baseline questionnaires. They were given contact details of the trial coordinator at baseline and at each follow-up assessment to ask any questions. Data were recorded in REDCap (Research Electronic Data Capture; hosted at the University of Queensland) [23,24]. In addition to primary and secondary outcomes, baseline data included duration of the current pain episode, demographics (age, gender, height, mass, education, job, and job status), and details about their low back symptoms (including location, intensity, duration, frequency, and past treatments; see Table 1 for specific questions and response options).

The participants were sent two reminders to complete the baseline data. Primary and secondary outcomes were collected at baseline and at 1, 3, 6, and 12 months after randomization. Pain intensity (pain visual analogue scale; see *Secondary Outcomes*), websites visited, and treatments used were recorded weekly until 3 months (primary endpoint) and then monthly until 12 months. Outcome data at each time point were collected using REDCap (automated email reminders), and up to three reminders were sent. To encourage retention, participants were entered into a draw for an iPad mini on completion of all data collection.

Table 1. Baseline characteristics of participants by group.

Characteristics	Group 1 (n=226)	Group 2 (n=214)
Acute[a], n (%)	59 (26.1)	53 (24.8)
Chronic[b], n (%)	167 (73.9)	161 (75.2)
Age at consent (years), mean (SD)	47.8 (14.1)	48.1 (14.0)
Gender, n (%)		
Female	174 (77)	161 (75.2)
Male	49 (21.7)	53 (24.8)
Other	3 (1.3)	0 (0)
Height (cm), mean (SD)	168.8 (10.1)	169.6 (9.3)
Mass (kg), mean (SD)	83.7 (20.1)	83.7 (20.7)
BMI (kg/m^2), median (IQR)	28.9 (24.5-33.6)	27.7 (24.1-32.6)
Level of education, n (%)		
High school certificate	38 (16.8)	30 (14)
Trade certificate	25 (11.1)	22 (10.3)
Diploma	37 (16.4)	29 (13.6)
Advanced diploma	12 (5.3)	13 (6.1)
Bachelor degree	59 (26.1)	57 (26.6)
Postgraduate degree	36 (15.9)	52 (24.3)
Other	19 (8.4)	11 (5.1)
Employment status, n (%)		
Full time/full duties	77 (34.1)	81 (37.9)
Full time/selected duties	8 (3.5)	2 (0.9)
Part time/full duties	33 (14.6)	43 (20.1)
Part time/selected duties	22 (9.7)	9 (4.2)
Not working/unemployed	4 (1.8)	6 (2.8)
Not working/employed/retraining	2 (0.9)	1 (0.5)
Not working/unemployed/retraining	4 (1.8)	7 (3.3)
Not working/unemployed	13 (5.8)	12 (5.6)
Not seeking employment	63 (27.9)	53 (24.8)
Aboriginal/Torres Strait Islander, n (%)	3 (1.3)	3 (1.4)
Born in Australia, n (%)	159 (70.4)	161 (75.2)
First experience of LBP[c], n (%)	14 (6.2)	12 (5.6)
Number of episodes of LBP, n (%)		
1-5	24 (11.3)	16 (8)
5-10	16 (7.5)	23 (11.4)
10-15	11 (5.2)	10 (5)
15-20	7 (3.3)	14 (7)
More than 20	63 (29.7)	54 (26.9)
I am never without LBP	91 (42.9)	84 (41.8)
Ever been given a diagnosis of LBP, n (%)	128 (56.6)	125 (58.4)
Have pain or altered sensation in buttocks/legs, n (%)	166 (73.5)	155 (72.4)
Had problems with your bowel or bladder function since your back pain started, n (%)		

Characteristics	Group 1 (n=226)	Group 2 (n=214)
Bladder	21 (9.3)	17 (7.9)
Bowel	22 (9.7)	15 (7)
Bladder and bowel	30 (13.3)	22 (10.3)
No	153 (67.7)	160 (74.8)
Had treatment for current episode of LBP, n (%)	142 (62.8)	116 (54.2)
Had treatments for previous episodes of LBP, n (%)	174 (82.5)	166 (82.6)
Have other medical conditions, n (%)	141 (62.4)	134 (62.6)
Have pain in other part of spine, n (%)	134 (59.3)	132 (61.7)
Feel pain in other areas of body, n (%)	143 (63.3)	130 (60.7)

[a] <6 weeks duration following period of no pain for at least 4 weeks.

[b] ≥6 weeks duration.

[c] LBP: low back pain.

Randomization and Allocation Concealment

The study biostatistician prepared the randomization schedule using stratification permuted block randomization, with block sizes of 6 to 12 stratified by symptom duration (acute LBP: <6 weeks duration with at least 4 weeks interval from a preceding episode; chronic LBP: all other presentations). After completion of baseline data, participants were randomized sequentially according to the schedule and provided with information regarding their mode of internet access.

Blinding

Participants and investigators (except the project manager) were blinded to treatment allocation. All participants were advised in the participant information sheet that the study aimed to investigate the impact of internet use on their LBP. To maintain blinding, all participants recorded the address of any website visited for information regarding LBP over 12 months. The intervention group had access to the password-protected MyBackPain website. The control group was not aware of this website and was unable to access it until the website was launched to the public on July 30, 2019, after the primary endpoint had passed for all participants. Given that all outcome measures were self-reported, outcome assessments were also blinded. A blinded biostatistician conducted the data analysis.

Study Treatments

Treatment

Participants randomized to the intervention group were given access to the MyBackPain website and encouraged to use it. The evidence-based content and framework of the MyBackPain website were developed according to consultation and collaboration with individuals with LBP, clinicians, and an international expert team [21].

Website development involved distillation of the highest quality information for acute and chronic LBP into easily understood resources in multiple formats. Textbox 1 lists the overarching principles.

Website access required a unique username and password provided to the intervention group participants. As we intended to study the natural use of the website, participants were free to determine how and when they accessed it and the content they used. Participants could use the website in multiple ways: self-directed browsing; automated, guided content tailored to the features of their presentation (evidence-based algorithms based on a pick-up tool for acute LBP [25] or STarT Back for chronic LBP [26]); and to *opt-in* for emails of key messages about LBP.

Textbox 1. Four overarching principles of the MyBackPain website.

Principles

- To enhance consumer confidence to manage their condition and make evidence-based treatment choices and avoid ineffective, unnecessary or harmful investigations and treatments
- To demedicalize and normalize low back pain (LBP) with messages that reinforce that back pain is a natural part of life for many and in most cases can be managed with early return to activity
- To provide tools for individuals to identify the necessity for further investigation, management, or both
- To engage users in healthy behaviors and attitudes to reduce the burden of LBP

Control

Participants randomized to the control group were asked to use the internet to obtain information about LBP in a self-directed manner and record the address of any relevant websites weekly (weeks 1-12) and monthly (months 3-12) web-based diaries. They did not have access to or knowledge of the MyBackPain website, at least until after the primary endpoint. We cannot exclude the possibility that some control participants might have become aware of and used the site after its launch. At launch, 89 control participants were yet to complete a long-term follow-up.

Outcome Measures

Primary Outcome

We used dimensions 2 and 3 of the Health Literacy Questionnaire (HLQ) [27] as coprimary outcomes to determine the extent to which participants considered "having sufficient information to manage my health" and "actively managing my health," respectively, at 3 months after randomization. The validated HLQ includes 44 items in nine dimensions, with each dimension considered individually [27]. Dimensions 2 and 3 included four items assessed using a 4-point Likert scale (1=strongly disagree; 4=strongly agree), converted to a score of 0 to 100 for analysis. The survey's preamble asked participants to consider their LBP when answering.

Secondary Outcomes

Secondary outcomes were (1) other HLQ dimensions, (2) quality of treatment choices, and (3) clinical outcomes. Secondary measures from the HLQ were dimensions: 1 ("feeling understood and supported by healthcare providers") and 4-9 ("social support for health;" "appraisal of health information;" "ability to actively engage with healthcare providers;" "navigating the healthcare system;" "ability to find good health information" and "understand health information well enough to know what to do"), assessed using 4-point (dimensions 1 and 4-5) or 5-point (dimensions 6-9) Likert scales.

The stated and observed quality of treatment choices was assessed in three ways:

1. Quality of treatment preference (stated): participants indicated on a 5-item scale (effective, somewhat effective, unsure, not very effective, and not effective) the degree to which they considered a subset of 10 treatments discussed in the MyBackPain website to be effective for LBP in response to the question "Do you think these treatments are effective for people's back pain? Note: think broadly about back pain, not just about your own." Treatment choices were scored against the recommendations provided in the MyBackPain website according to the classifications of "good evidence," "may work," "not enough evidence," "unlikely to work," and "may be harmful" (Multimedia Appendix 1 shows the scoring matrix; scores from −20 to +22; a 1-point change would relate to a shift from being *unsure* of the efficacy of one treatment that is *unlikely to work* to considering it to be *not very effective*).
2. Quality of treatment preference (observed-scored): treatments used by participants (diary recording) were evaluated against the recommendations provided on the MyBackPain website. Each reported treatment was scored according to Multimedia Appendix 1 and summed (no upper and lower limits).
3. Quality of treatment preference (observed-proportion): the proportion of participants who chose treatments that are, according to MyBackPain, either recommended ("good evidence" and "may work") or considered to have no effect or be harmful ("not enough evidence," "unlikely to work," and "may be harmful") were assessed separately. Using diary data, participants were allocated a score of 1 if they used at least one recommended treatment and separately if they used a treatment that was harmful or had no effect. Proportions with scores of 1 were calculated.

LBP clinical outcomes were measured with the following validated tools: pain—visual analogue scale of average LBP in the last week ("no pain" at 0 and "worst pain imaginable" at 100); disability—Roland Morris Disability Questionnaire [28] (scores from 0 to 24; higher scores indicate more disability); and quality of life—Assessment of Quality of Life–8 dimension (utility scores from 0.0 to 1.0, with higher scores indicating better quality of life).

Adherence to Intervention

We were unable to track the individual use of MyBackPain because of privacy concerns of the host sites (Arthritis Australia). We tracked the overall use of the site (number of sessions; new or returning users) using the Opentracker software.

Sample Size

Sample size calculation was based on an effect size of 0.30 for the coprimary outcomes. A sample size of 440 participants (a minimum of 110/440, 25% with acute LBP) was required to achieve 80% power to detect the desired effect size, type 1 error of 0.05, and allowing for a 20% loss to follow-up at 3 months.

Statistical Analysis

Data were analyzed using Stata version 16.0 (StataCorp) and included all participants in their randomized groups (intention-to-treat). The baseline characteristics of participants who did and did not provide primary outcomes were compared using two-tailed *t* tests or chi-square tests. Missing continuous outcomes were imputed using chained equations with predictive mean matching and five nearest neighbors, and missing binary outcomes were imputed using logistic regression imputation models with chained equations. The data were imputed separately for each randomized group. Imputation models for continuous outcomes included all baseline variables and all outcome variables, where possible. Imputation models for binary variables omitted binary baseline variables and outcome variables because of the potential for perfect prediction, including only continuous baseline variables. Estimates from 40 imputed data sets were combined using the Rubin rules [29].

For the primary outcomes and continuous secondary outcomes, mean differences between groups were estimated at each time point using longitudinal linear mixed-effects models. These models included all data from 1, 3 (primary endpoint), 6, and 12 months as outcomes for each participant. The models included an interaction between month and randomized group as a fixed effect and random intercepts for participants. Models were adjusted for baseline values of the outcome and stratification variables (symptom duration). Similar longitudinal logistic regression models were used for secondary binary outcomes, with results presented as odds ratios and risk ratios at each time point. Model assumptions and the validity of the imputed data sets were assessed using standard diagnostic plots. No statistical adjustments were made for multiple testing. Complete case analyses were also performed.

Results

Recruitment was conducted from December 6, 2017, to January 16, 2019, with follow-up completed on January 16, 2020. We enrolled 453 participants who completed the baseline assessment (Figure 1). In total, 13 withdrew (different time points), and their data were excluded from the analysis. The groups were similar at baseline (Multimedia Appendix 2). Loss to follow-up at the primary endpoint (3 months) was 26.1% (59/226) in the control group and 28% (60/214) in the MyBackPain group. Those lost to follow-up were younger and had fewer previous episodes of LBP (Multimedia Appendix 2).

For the primary outcomes at 3 months, between-group differences in dimensions 2 (mean difference −0.87 units, 95% CI −3.56 to 1.82) and 3 (mean difference −0.41 units, 95% CI −2.78 to 1.96) of the HLQ were small with wide CIs (Multimedia Appendix 3).

For secondary outcomes, the MyBackPain group scored greater for treatment choice (stated) than the control group at 1 month (−0.91 units, 95% CI 0.16 to 1.67), but differences beyond that time point were of inconsistent direction and unlikely to be clinically significant (Multimedia Appendix 3), as was the observed proportion of participants selecting a recommended treatment at all time points (Table 2). Between-group differences in other secondary outcomes (other HLQ dimensions and clinical outcomes) at any time point were of inconsistent direction.

The results for the complete case analysis were similar (Multimedia Appendices 4 and 5), except that more participants (observed-proportion) in the treatment group were likely to select a recommended treatment (odds ratio 5.76, 95% CI 1.01 to 32.71; risk ratio 1.29, 95% CI 1.01 to 1.58) at 3 months (Table 2).

Up to the launch date (after the primary endpoint for all participants), the MyBackPain website was accessed an average of 3.4 times, each by 183 unique users. This represented 85.5% (183/214) of the patients allocated to the MyBackPain group.

Figure 1. CONSORT (Consolidated Standards of Reporting Trails) flow diagram. LBP: low back pain.

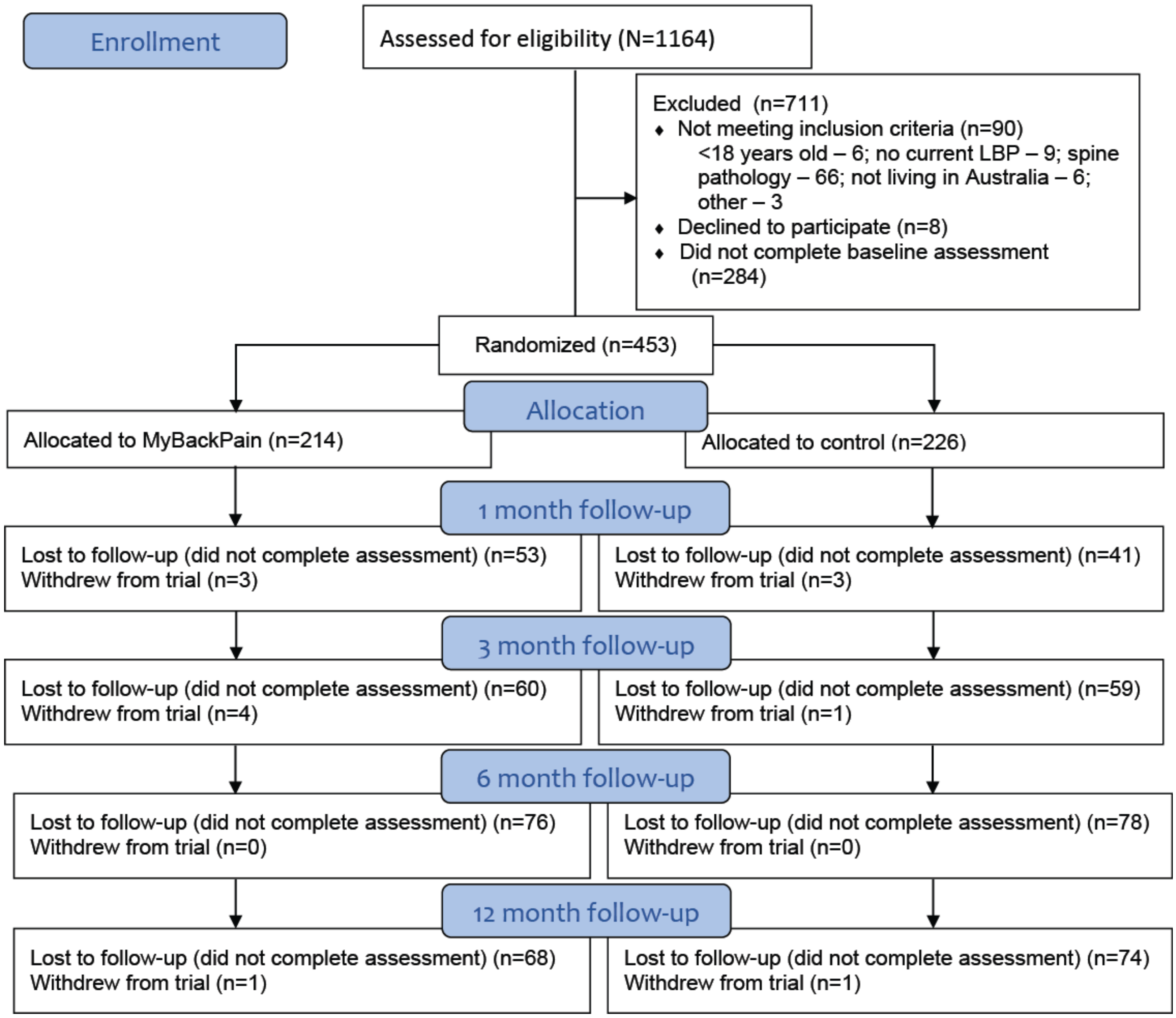


Table 2. Participants choosing treatment types (the binary outcomes)[a].

Treatment type and month	Control			MyBackPain			Risk ratio (95% CI)	*P* value
	Total, participant, n	Participant, n (%)	Number missing (n=226), n (%)	Total, participant, n	Participant, n (%)	Number missing (n=214), n (%)		
Observed treatment choice harmful or no effect								
1	83	20 (24.1)	143 (63.2)	63	17 (26.9)	151 (70.6)	1.07 (0.55-1.59)	.79
3	73	12 (16.4)	153 (67.7)	52	13 (25)	162 (75.7)	1.39 (0.56-2.23)	.35
6	67	11 (16.4)	159 (70.3)	67	11 (16.4)	147 (68.7)	0.76 (0.30-1.22)	.30
12	89	18 (20.2)	137 (60.6)	72	10 (13.8)	142 (66.3)	0.94 (0.40-1.48)	.83
Observed treatment choice recommended								
1	83	55 (66.2)	143 (63.2)	63	40 (63.5)	151 (70.6)	0.89 (0.69-1.08)	.26
3	73	49 (67.1)	153 (67.7)	52	37 (71.1)	162 (75.7)	0.99 (0.78-1.20)	.94
6	67	54 (80.6)	159 (70.3)	67	47 (70.1)	147 (68.7)	0.84 (0.64-1.05)	.13
12	89	65 (73)	137 (60.6)	72	54 (75)	142 (66.3)	0.99 (0.78-1.20)	.91

[a] Odds and risk ratios are calculated using multiply imputed data. Counts and percentages calculated using the observed data only.

Discussion

Principal Findings

The results of this study showed that the natural use of a web-based information-only resource (MyBackPain [20]) did not improve the two dimensions of health literacy more than self-guided access to web-based LBP information. Use of the website did not achieve better outcomes for other secondary measures of clinical features of pain, disability, quality of life, or other aspects of health literacy, except for the stated and observed selection of recommended treatments at one follow-up time point.

Interpretation of Findings

MyBackPain is a website for individuals with LBP that was developed with extensive consumer input to ensure that content and design are aligned with patient needs and preferences [21]. The website's intention is to enhance an individual's capacity or confidence to manage their condition, including decision support for the selection of effective treatments. The primary outcomes for this randomized controlled trial were selected to reflect this goal with respect to "having sufficient information to manage my health" or "actively managing my health." Contrary to our hypothesis, the use of the MyBackPain website did not achieve greater improvement in these domains than the unguided use of the internet. However, the finding that access to MyBackPain achieved greater improvement in stated treatment choices at 1 month provides preliminary evidence for the improvement of active management of their condition.

Health literacy is defined 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" [30]. Health literacy extends beyond the provision of information to the empowerment of individuals to use information [27]. The use of MyBackPain did not more effectively modify an individual's *perception* of their ability to manage their own health than unguided use of the internet, but the data suggest some short-term improvement in treatment choices. One interpretation is that improvement in these elements of health literacy may be satisfied by the general use of information on the internet but that the quality of the resulting actions (ie, treatment selection) depends on the quality of the information provided. The control group's lower treatment choice scores concur with a recent review that identified only 43.3% of treatment recommendations provided by freely accessible websites were accurate [17]. The treatment recommendations provided in MyBackPain are based on expert panel consensus.

On the basis of population data for the HLQ domains from the Australian Bureau of Statistics, our sample had similar health literacy at baseline (dimension 2: our data=mean 3.12, SD 0.6, vs Australian Bureau of Statistics data for individuals who report a *back problem*=mean 3.13, relative SE 0.5; dimension 3: 3.23, SD 0.6, vs 3.03, SE 0.6). However, our sample had low quality of life utility scores compared with Australian norms (our data: mean 0.55, SD 0.22, vs mean 0.799, 95% CI 0.792-0.806 [31]). Despite the co-design of MyBackPain in individuals with LBP, the presentation of information might not be optimal for those with low quality of life [32].

MyBackPain was designed as an information resource. The information is provided in multiple formats but with limited interactivity. Benefits might be enhanced by greater interactivity, such as inclusion of a web-based discussion group [33], regular action plans [34], or provision of active intervention. Refinement of MyBackPain would require assessment because recent evidence highlights inconsistent evidence for the efficacy of web-based cognitive behavioral therapies for LBP [18,19,35,36], interactivity [12], and personalization [36,37].

Treatment Selection

This study provides preliminary evidence that access to MyBackPain might achieve short-term improvement in treatment choices. The largest component of the website, as requested by consumers [10], is a treatment comparison tool. This tool provides summaries of the treatment, the quality of evidence for efficacy, and pros and cons for its use with an evidence icon that provides a quick reference of the level of evidence and effectiveness [21]. Summaries were developed with input from an international expert team and consumers [21]. Although no other studies have tested the effectiveness of web-based information to improve treatment choices, studies that are not specific to LBP have reported reduced health care usage (improved self-management and decreased use of ineffective treatments) related to high-quality, evidence-based, web-based information [38].

We did not include treatment selection as a primary outcome because of the novelty of our measures. The measures were tested for comprehension in individuals with LBP but had unknown psychometric properties. The *observed* and *stated* measures of treatment choice were based on expert consensus recommendations in the website. Alignment of stated and observed choices with the recommended treatments resulted in a higher score. Better stated treatment choices by the MyBackPain group at 1 month provide stimulus for further work. No group difference at 3 months implies that sustained improvement might require sustained engagement with the website.

Web Resource Evaluation

It is recommended that information resources be tested for the credibility and accuracy of content [39]. According to these criteria, the MyBackPain website has a strong foundation supported by research with consumers [32] and expert consultation for content accuracy [21,40]. Multimedia Appendix 6 [10,21,25,32,41-43] considers MyBackPain against evidence standards for digital health technologies [41].

Limitations

There are several limitations that need to be considered. First, there was a large loss to follow-up, which is common in web-based trials [35,44]. Although multiple imputations were implemented, if participants were lost to follow-up due to worsening of their condition, this analysis may not be appropriate. Second, the individual use of MyBackPain could not be tracked. That 14.5% (31/214) of the treatment group did not access MyBackPain is disappointing, but it is important to note that the trial was pragmatic and designed to evaluate natural use.

Conclusions

This is the first randomized controlled trial of an information resource for LBP to test its impact on health literacy, quality of treatment choices, and clinical outcomes. Unlike most websites for LBP, MyBackPain was specifically designed to meet the needs of individuals with LBP and included a focus on treatment choices. The results provided no evidence of differences between groups for the primary outcomes related to health literacy. There is limited preliminary evidence of the impact of empowerment on making informed choices for the management of their condition, which is the intention of improved health literacy. Loss to follow-up should be considered when interpreting these results. Future work should consider the potential to enhance impact through the addition of interactivity and interventions and greater support provided to users to engage with the resource.

Acknowledgments

This work was supported by grants from the National Health and Medical Research Council (NHMRC, Australia; APP1079078 and APP1091302) and the Medibank Health Research Fund. PWH (APP1102905), MF (APP1143593), DH (APP1194737), KB (APP1174431), and JS (APP1157179) are supported by fellowships from the NHMRC. The NHMRC (Australia) and the Medibank Health Research Fund had no role in the study design, collection, analysis, interpretation of data, or the decision to publish.

Authors' Contributions

PWH had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors contributed to the conceptualization and design of the study, data acquisition, analysis and interpretation, and drafting of the manuscript. All authors critically revised the manuscript for important intellectual content. Statistical analysis was primarily performed by JK and SC, and funding was obtained by PWH, KB, DH, and BV.

Conflicts of Interest

DH is on advisory boards for Merck Serono, Tissuegene, TLC Bio, Pfizer, Lilly, Novartis, Biobone, and Flexion, outside of the submitted work.

Multimedia Appendix 1
Scoring matrix for evaluation of quality of treatment preference.
[DOCX File , 21 KB - jmir_v23i6e27860_app1.docx]

Multimedia Appendix 2
Baseline characteristics and outcome scores of participants who did and did not report both primary outcomes, reported as number (percentage) unless otherwise stated.

[DOCX File , 27 KB - jmir_v23i6e27860_app2.docx]

Multimedia Appendix 3
Mean (SD) of continuous outcomes and mean difference (95% CI) between MyBackPain and control groups over the 12-month study period.
[DOCX File , 23 KB - jmir_v23i6e27860_app3.docx]

Multimedia Appendix 4
Complete case data: mean (SD) of continuous outcomes and mean difference between groups over time (MyBackPain and control).
[DOCX File , 30 KB - jmir_v23i6e27860_app4.docx]

Multimedia Appendix 5
Complete case data: number (percentage) of participants choosing treatment types (binary outcomes).
[DOCX File , 22 KB - jmir_v23i6e27860_app5.docx]

Multimedia Appendix 6
Evidence for effectiveness standards.
[DOCX File , 21 KB - jmir_v23i6e27860_app6.docx]

Multimedia Appendix 7
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 385 KB - jmir_v23i6e27860_app7.pdf]

References

1. GBD 2017 Disease Injury Incidence Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Dec 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Buchbinder R, van Tulder M, Öberg B, Costa LM, Woolf A, Schoene M, Lancet Low Back Pain Series Working Group. Low back pain: a call for action. Lancet 2018 Jun 09;391(10137):2384-2388. [doi: 10.1016/S0140-6736(18)30488-4] [Medline: 29573871]
3. Alyousef B, Cicuttini FM, Davis SR, Bell R, Botlero R, Urquhart DM. Negative beliefs about back pain are associated with persistent, high levels of low back disability in community-based women. Menopause 2018 Sep;25(9):977-984. [doi: 10.1097/GME.0000000000001145] [Medline: 29870470]
4. Carey M, Turon H, Goergen S, Sanson-Fisher R, Yoong SL, Jones K. Patients' experiences of the management of lower back pain in general practice: use of diagnostic imaging, medication and provision of self-management advice. Aust J Prim Health 2015;21(3):342-346. [doi: 10.1071/PY14057] [Medline: 25074025]
5. Clarke MA, Moore JL, Steege LM, Koopman RJ, Belden JL, Canfield SM, et al. Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: a literature review. Health Informatics J 2016 Dec 15;22(4):992-1016 [FREE Full text] [doi: 10.1177/1460458215602939] [Medline: 26377952]
6. Kummervold PE, Chronaki CE, Lausen B, Prokosch H, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
7. Fox S. The social life of health information. Washington DC.: Pew research Center; 2009. URL: https://www.pewresearch.org/internet/2009/06/11/the-social-life-of-health-information/ [accessed 2020-10-09]
8. Volkman JE, Luger TM, Harvey KL, Hogan TP, Shimada SL, Amante D, et al. The National Cancer Institute's Health Information National Trends Survey [HINTS]: a national cross-sectional analysis of talking to your doctor and other healthcare providers for health information. BMC Fam Pract 2014;15:111 [FREE Full text] [doi: 10.1186/1471-2296-15-111] [Medline: 24906558]
9. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
10. Nielsen M, Jull G, Hodges PW. Information needs of people with low back pain for an online resource: a qualitative study of consumer views. Disabil Rehabil 2014;36(13):1085-1091. [doi: 10.3109/09638288.2013.829532] [Medline: 24001260]
11. Li N, Orrange S, Kravitz RL, Bell RA. Reasons for and predictors of patients' online health information seeking following a medical appointment. Fam Pract 2014 Oct;31(5):550-556. [doi: 10.1093/fampra/cmu034] [Medline: 24963151]

12. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 5;22(6):e15617. [doi: 10.2196/15617]
13. Butler L, Foster NE. Back pain online: a cross-sectional survey of the quality of web-based information on low back pain. Spine (Phila Pa 1976) 2003 Feb 15;28(4):395-401. [doi: 10.1097/01.BRS.0000048497.38319.D3] [Medline: 12590218]
14. Hendrick PA, Ahmed OH, Bankier SS, Chan TJ, Crawford SA, Ryder CR, et al. Acute low back pain information online: an evaluation of quality, content accuracy and readability of related websites. Man Ther 2012 Aug;17(4):318-324. [doi: 10.1016/j.math.2012.02.019] [Medline: 22464886]
15. Li L, Irvin E, Guzmán J, Bombardier C. Surfing for back pain patients: the nature and quality of back pain information on the Internet. Spine (Phila Pa 1976) 2001 Mar 01;26(5):545-557. [doi: 10.1097/00007632-200103010-00020] [Medline: 11242383]
16. Costa N, Nielsen M, Jull G, Claus AP, Hodges PW. Low back pain websites do not meet the needs of consumers: a study of online resources at three time points. Health Inf Manag 2019 Jul 02:A. [doi: 10.1177/1833358319857354] [Medline: 31266366]
17. Ferreira G, Traeger AC, Machado G, O'Keeffe M, Maher CG. Credibility, accuracy, and comprehensiveness of internet-based information about low back pain: a systematic review. J Med Internet Res 2019 May 07;21(5):e13357 [FREE Full text] [doi: 10.2196/13357] [Medline: 31066689]
18. Buhrman M, Nilsson-Ihrfeldt E, Jannert M, Ström L, Andersson G. Guided internet-based cognitive behavioural treatment for chronic back pain reduces pain catastrophizing: a randomized controlled trial. J Rehabil Med 2011 May;43(6):500-505 [FREE Full text] [doi: 10.2340/16501977-0805] [Medline: 21533329]
19. Carpenter KM, Stoner SA, Mundt JM, Stoelb B. An online self-help CBT intervention for chronic lower back pain. Clin J Pain 2012 Jan;28(1):14-22 [FREE Full text] [doi: 10.1097/AJP.0b013e31822363db] [Medline: 21681084]
20. MyBackPain. URL: http://mybackpain.org.au [accessed 2020-10-07]
21. Hodges PW, Setchell J, Nielsen M. An internet-based consumer resource for people with low back pain (MyBackPain): development and evaluation. JMIR Rehabil Assist Technol 2020 Mar 31;7(1):e16101 [FREE Full text] [doi: 10.2196/16101] [Medline: 32229467]
22. Hall LM, Ferreira M, Setchell J, French S, Kasza J, Bennell KL, et al. MyBackPain-evaluation of an innovative consumer-focused website for low back pain: study protocol for a randomised controlled trial. BMJ Open 2019 May 14;9(5):e027516 [FREE Full text] [doi: 10.1136/bmjopen-2018-027516] [Medline: 31092664]
23. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
24. 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]
25. Traeger AC, Henschke N, Hübscher M, Williams CM, Kamper SJ, Maher CG, et al. Estimating the risk of chronic pain: development and validation of a prognostic model (PICKUP) for patients with acute low back pain. PLoS Med 2016 May;13(5):e1002019 [FREE Full text] [doi: 10.1371/journal.pmed.1002019] [Medline: 27187782]
26. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis Rheum 2008 May 15;59(5):632-641 [FREE Full text] [doi: 10.1002/art.23563] [Medline: 18438893]
27. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
28. Roland M, Morris R. A study of the natural history of low-back pain. Part II: development of guidelines for trials of treatment in primary care. Spine (Phila Pa 1976) 1983 Mar;8(2):145-150. [doi: 10.1097/00007632-198303000-00005] [Medline: 6222487]
29. Carpenter J, Kenward M. Multiple Imputation and its Application, 1st Edition. West Sussex, UK: John Wiley & Sons Ltd; 2013:1-366.
30. Health promotion glossary. World Health Organization. 1998. URL: https://www.who.int/healthpromotion/about/HPR%20Glossary%201998.pdf?ua=1 [accessed 2021-06-04]
31. Maxwell A, Özmen M, Iezzi A, Richardson J. Deriving population norms for the AQoL-6D and AQoL-8D multi-attribute utility instruments from web-based data. Qual Life Res 2016 Dec;25(12):3209-3219. [doi: 10.1007/s11136-016-1337-z] [Medline: 27344318]
32. Setchell J, Turpin M, Costa N, Hodges P. Web-based consumer health education about back pain: findings of potential tensions from a photo-elicitation and observational study. JMIR Rehabil Assist Technol 2020 Jun 01;7(1):e17130 [FREE Full text] [doi: 10.2196/17130] [Medline: 32478663]
33. Maclachlan LR, Mills K, Lawford BJ, Egerton T, Setchell J, Hall LM, et al. Design, delivery, maintenance, and outcomes of peer-to-peer online support groups for people with chronic musculoskeletal disorders: systematic review. J Med Internet Res 2020 Apr 24;22(4):e15822 [FREE Full text] [doi: 10.2196/15822] [Medline: 32329746]

34. Riva S, Camerini A, Allam A, Schulz PJ. Interactive sections of an internet-based intervention increase empowerment of chronic back pain patients: randomized controlled trial. J Med Internet Res 2014;16(8):e180 [FREE Full text] [doi: 10.2196/jmir.3474] [Medline: 25119374]
35. Garg S, Garg D, Turin TC, Chowdhury MF. Web-based interventions for chronic back pain: a systematic review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
36. Irvine AB, Russell H, Manocchia M, Mino DE, Cox GT, Morgan R, et al. Mobile-Web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
37. Lustria ML, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
38. Spoelman WA, Bonten TN, de Waal MW, Drenthen T, Smeele IJ, Nielen MM, et al. Effect of an evidence-based website on healthcare usage: an interrupted time-series study. BMJ Open 2016 Dec 09;6(11):e013166 [FREE Full text] [doi: 10.1136/bmjopen-2016-013166] [Medline: 28186945]
39. Greaves F, Joshi I, Campbell M, Roberts S, Patel N, Powell J. What is an appropriate level of evidence for a digital health intervention? Lancet 2019 Dec 22;392(10165):2665-2667. [doi: 10.1016/S0140-6736(18)33129-5] [Medline: 30545779]
40. French SD, Nielsen M, Hall LM, Nicolson PJ, van Tulder M, Bennell KL, et al. Essential key messages about diagnosis, imaging, and self-care for people with low back pain: a modified Delphi study of consumer and expert opinions. Pain 2019 Dec 23;160(12):2787-2797. [doi: 10.1097/j.pain.0000000000001663] [Medline: 31356451]
41. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. 2018. URL: https://www.nice.org.uk/corporate/ecd7 [accessed 2020-10-07]
42. Nielsen M, Jull G, Hodges PW. Designing an online resource for people with low back pain: health-care provider perspectives. Aust J Prim Health 2016;22(2):159-166. [doi: 10.1071/PY14131] [Medline: 25705821]
43. Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet 2011 Oct 29;378(9802):1560-1571 [FREE Full text] [doi: 10.1016/S0140-6736(11)60937-9] [Medline: 21963002]
44. Weymann N, Dirmaier J, von Wolff A, Kriston L, Härter M. Effectiveness of a web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]

Abbreviations

HLQ: Health Literacy Questionnaire
LBP: low back pain
REDCap: Research Electronic Data Capture

Edited by R Kukafka; submitted 10.02.21; peer-reviewed by J Hartvigsen, C Maher; comments to author 16.03.21; revised version received 25.03.21; accepted 06.05.21; published 15.06.21.

Please cite as:
Hodges PW, Hall L, Setchell J, French S, Kasza J, Bennell K, Hunter D, Vicenzino B, Crofts S, Dickson C, Ferreira M
Effect of a Consumer-Focused Website for Low Back Pain on Health Literacy, Treatment Choices, and Clinical Outcomes: Randomized Controlled Trial
J Med Internet Res 2021;23(6):e27860
URL: https://www.jmir.org/2021/6/e27860
doi:10.2196/27860
PMID:34128822

Original Paper

Rating Hospital Performance in China: Review of Publicly Available Measures and Development of a Ranking System

Shengjie Dong[1*], MPH; Ross Millar[2*], PhD; Chenshu Shi[3], MSc; Minye Dong[1], MPH; Yuyin Xiao[1], MPH; Jie Shen[4], MD; Guohong Li[1,4], PhD

[1]School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2]Health Services Management Centre, University of Birmingham, Birmingham, United Kingdom
[3]Center for Health Technology Assessment, China Hospital Development Institute, Shanghai Jiao Tong University, Shanghai, China
[4]China Hospital Development Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[*]these authors contributed equally

Corresponding Author:
Guohong Li, PhD
China Hospital Development Institute
Shanghai Jiao Tong University School of Medicine
227 South Chong Qing Road
Shanghai, 200025
China
Phone: 86 21 63846590
Email: guohongli@sjtu.edu.cn

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/6/e31370

Abstract

Background: In China, significant emphasis and investment in health care reform since 2009 has brought with it increasing scrutiny of its public hospitals. Calls for greater accountability in the quality of hospital care have led to increasing attention toward performance measurement and the development of hospital ratings. Despite such interest, there has yet to be a comprehensive analysis of what performance information is publicly available to understand the performance of hospitals in China.

Objective: This study aims to review the publicly available performance information about hospitals in China to assess options for ranking hospital performance.

Methods: A review was undertaken to identify performance measures based on publicly available data. Following several rounds of expert consultation regarding the utility of these measures, we clustered the available options into three key areas: research and development, academic reputation, and quality and safety. Following the identification and clustering of the available performance measures, we set out to translate these into a practical performance ranking system to assess variation in hospital performance.

Results: A new hospital ranking system termed the China Hospital Development Index (CHDI) is thus presented. Furthermore, we used CHDI for ranking well-known tertiary hospitals in China.

Conclusions: Despite notable limitations, our assessment of available measures and the development of a new ranking system break new ground in understanding hospital performance in China. In doing so, CHDI has the potential to contribute to wider discussions and debates about assessing hospital performance across global health care systems.

(J Med Internet Res 2021;23(6):e17095) doi:10.2196/17095

KEYWORDS

hospital ranking; performance measurement; health care quality; China health care reform

Introduction

Hospital rating systems have the potential to play an important role in patient decision-making as well as offer policy makers and practitioners valuable opportunities to monitor and improve the quality of hospital services [1-4]. In China, significant emphasis and investment into health care reform since 2009 has brought with it increasing scrutiny of its public hospitals with regard to improving their quality and efficiency. Reform measures have included an emphasis on improving hospital governance with clearer regulations and transparency regarding overall performance [5]. Although these measures show promising signs, questions remain about their overall impact and sustainability [6], as well as those concerning the information asymmetries that exist for patients and providers that limit the market conditions of competition and choice deemed necessary to rate hospital performance [7].

An enduring feature of China's health care provision is the dominance of the hospital sector. Within these contexts, patients are offered different forms of provision ranging from grade I community hospitals, grade II secondary or county hospitals serving several communities, and grade III tertiary hospitals serving districts or cities. This classification [8] remains a powerful driving force for decision-making, with tertiary hospitals often deemed the preferred option for better clinical quality. Pan et al [7] explain how such trends are driven by a culture where patient volume often represents the primary measure of hospital performance used by government administrators. Patients often equate hospital size as a signal of quality, thus preferring to self-refer to larger tertiary hospitals. Large patient volume is also deemed essential for hospitals in developing a good reputation and acquiring high-quality research and training programs.

U.S. News & World Report's Best Hospitals ranking is one of the well-known hospitals ranking systems that aims to help patients find professional medical centers and doctors across the United States. The relative success of the Best Hospitals ranking demonstrates that the objectivity of measures such as mortality and morbidity can provide an important contribution for accurate evaluation of health care quality [9]. However, influential rankings in developed countries, such as Best Hospitals ranking and Vizient Award [10], are based on solid medical information supporting mechanisms and are challenging to be applied to low- and middle-income countries or regions with relatively underdeveloped medical information supporting facilities.

In order to try and disentangle these trends, China is increasingly seeing the development and use of hospital performance rankings. The annual publication of the Hospital Management Institute of Fudan University Hospital Ranking list [11] ranks hospitals according to a social reputation score that is determined based on survey responses from physicians combined with a review of scientific research outputs from their affiliated institutions. The Science and Technology Evaluation Metrics (STEM) of hospitals developed by the Chinese Academy of Medical Sciences [12] ranks tertiary hospital performance based on their science and technology investment and any associated outputs. Other influential rankings include the top-100 China Hospitals Competitiveness by Alibi Hospital Management Research Center, Hong Kong [13], and China's best clinical discipline rankings released by Peking University [14].

These indicators provide a valuable contribution to debates and decision-making about hospital performance in China. Nevertheless, given the current situation in China, and the asymmetries of information that exist, important limitations of the current ranking systems have been highlighted, including a reliance on reputation scores [11], a limited menu of performance measures, and a lack of consideration and engagement with measures of quality and safety [8].

Current ranking systems undoubtedly have merit in signifying attempts to better understand hospital performance; however, further research is needed to better understand and triangulate publicly available hospital performance information. Thus far, there has yet to be a comprehensive analysis of what performance information is actually available in China [15]. This study aims to review the publicly available performance information with the view to assess different ways in which hospital performance can be ranked. In doing so, in this paper, we present a new hospital ranking system, termed the "China Hospital Development Index" (CHDI). Although this ranking system faces notable limitations, we argue that our review of measures and development of a ranking system break new ground that can inform both current and future policy and practice for hospital performance in China.

Methods

Limitations of Major Hospital Ranking Systems in China

Current hospital performance rankings in China [11-13] classify hospital performance across a range of indictors, including the availability of hospital facilities, services, and personnel; the calculation of social reputation scores; and the publication of scientific research inputs and outputs. These indicators provide valuable contributions for understanding hospital performance; however, a notable limitation in the rankings produced so far has been the emphasis on the quality and safety of health care provision. Quality and safety represent core domains of medical services; therefore, any assessment of hospital performance should aim to incorporate any available measures [16].

It is worth pointing out that for the clinical disciplines ranking reported in Table 1, more than 48 million clinical data records were collected from nearly 400 hospitals across China from 2006 to 2014. The main characteristic of this ranking is that the focus has been shifted to the clinical specialties rather than the number of funds and articles published; it is also the first application of effective medical clinical data for evaluation of hospitals. It should be mentioned that its methodology has not been made public. However, this ranking was unsuccessful (published only once in 2015). The main reason is the standardization of clinical data, such as inconsistent disease codes, which directly affects the quality of medical record information used. Although hospitals in China are vigorously promoting medical informatization at this stage, there is still a long way to use medical data to rank hospitals even if such data are available.

Table 1. Overview of China's hospital ranking systems.

Characteristic	Best Hospitals ranking [11]	STEM[a] [12]	Hospital competitiveness ranking [13]	Best Clinical Disciplines ranking [14]
Primary objective	To provide guidelines for patients seeking treatments	To measure a hospital's value in scientific research	To identify the top hospitals with the best competitiveness	To provide a tool to help patients find skilled specialty care
Domains of measure	• Social reputation • The ability of sustainable development (scientific research outcomes)	• Scientific research inputs, outputs, and impacts	• Medical service • Academic impacts • Resource management • Hospital operation	Unclear
Indicators	• Social reputation scores • SCI[b] papers • National awards	• Key laboratories and projects • Researchers • Clinical trials • SCI papers • Medical standards • Medical association leaders • National awards	• Inpatients and outpatients • Beds • Health workers • Medical facilities • Personnel • Medical fee • Length of stay • Academic leaders • Key laboratories and projects • National awards	• Perioperative mortalities • Readmissions • Postoperative complications • Timely care • Finance
Data sources	National surveys	SCI database and official documents	Reporting data voluntarily	Medical records
Publication frequency	Annually	Annually	Annually	Only in 2015
Transparency of methodology	Provided	Provided	Provided	Not provided

[a] STEM: Science and Technology Evaluation Metrics.

[b] SCI: Science Citation Index.

Exploring Available Performance Measures

Based on the assessment of current measures and the identification of areas for improvement, a group of 6 experts with physician and methodological expertise in performance measurement was established within the China Hospital Development Institution (HDI) to assess the available options and provide feedback at each step of the process (see Multimedia Appendix 1). To begin the analysis, we mapped publicly available measures and identified different information sources that reflected our interest in better understanding *hospital performance*. Through iterative discussions among the study group, a review of the available literature, and discussions with experts, we established three performance domains for the purpose of ranking hospitals in China. These domains were categorized as research and development, academic reputation, and quality and safety (described below and summarized in Table 2).

Table 2. Summary of measures and clustering of performance indicators into three domains: research and development, academic reputation, and quality and safety.

Domains and measures	Indicators	Data sources
Research and development		
Publication outputs	Number of SCI[a] papers by authors, by first author, and by corresponding author	Web of Science [17]
Number of citations	Number of citations by authors, by first author, and by corresponding author	Web of Science [17]
Number of high-impact outputs	Number of SCI papers with journal IF[b] ≥10 by authors, by first author, and by corresponding author; number of papers published in top-6 journals by authors, by first author, and by corresponding author	Web of Science [17]
Clinical trial activity	Number of registered clinical trials	ChiCTR[c] [18]
Academic reputation		
Academician	Number of academicians of CAS[d] and CAE[e]	CAS [19], CAE [20]
Chief editor	Number of staff as chief editors of core medical journals included in CSCD[f]	CSCD, Science China [21]
Association chairperson or member	Number of staff as National Association chairperson and National Association members	CMA[g] [22], CMDA[h] [23]
Award	SPSTA[i], SNSA[j], STTPA[k], CMSTA[l], CDA[m]	CMA [22], CMDA [23], NOSTA[n] [24]
Quality and Safety		
Quality of specialty care	Number of national key clinical specialties, diagnosis and treatment, Improvement of Rare Diseases Program	Official websites of each hospital; NHC[o], People's Republic of China [25]
Medical malpractice claims	Ratio of compensation cases, ratio of liability	Laws and Regulations – Peking University [26]

[a] SCI: Science Citation Index.

[b] IF: impact factor.

[c] ChiCTR: Chinese Clinical Trial Registry.

[d] CAS: Chinese Academy of Sciences.

[e] CAE: Chinese Academy of Engineering.

[f] CSCD: China Science Citation Database.

[g] CMA: Chinese Medical Association.

[h] CMDA: Chinese Medical Doctor Association.

[i] SPSTA: State Preeminent Science and Technology Award.

[j] SNSA: State Natural Science Award.

[k] STTPA: State Scientific and Technological Progress Award.

[l] CMSTA: Chinese Medical Science and Technology Award.

[m] CDA: Chinese Doctor Award.

[n] NOSTA: National Office for Science and Technology Awards

[o] NHC: National Health Commission.

To begin our analysis of available measures, we used the Health Statistics Yearbook issued by the National Health Committee to gather baseline information regarding outpatient, inpatient, and emergency admissions to hospitals in China [27]. Second, given the importance placed on research and development in China as a measure of performance, we also sought to identify research and development indicators for hospitals gathered from research databases in order to gauge the research activity and outputs being produced by each hospital. This would include any hospital affiliation of authorship to published Science Citation Index (SCI) papers, the number of citations obtained, and the number of SCI papers for which an impact factor (IF) ≥10 per hospital. Information regarding clinical trial activity was also collected as an indicator for research activity.

As a follow-up to the research and development activity, we were able to obtain measures demonstrating clinical academic reputation of hospital staff engaged in high-impact research outputs demonstrating wide scholarly impact in their clinical area of expertise. This included measuring the number of academic affiliations with the Chinese Academy of Science (CAS) and the Chinese Academy of Engineering (CAE); the number of staff as chief editors of core medical journals included in the China Science Citation Database (CSCD); membership of national associations, including the Chinese Medical Association (CMA) and the Chinese Medical Doctor Association (CMDA); and the number of national awards received per hospital, including the State Preeminent Science and Technology Award (SPSTA), the State Natural Science Award (SNSA), the State Scientific and Technological Progress Award (STTPA), the Chinese Medical Science and Technology Award (CMSTA), and the Chinese Doctor Award (CDA).

Finally, our analysis of the quality and safety performance measures was able to draw on medical malpractice litigation records [28,29] adjusted for complexity and risk of patient disease as two useful indicators for patient safety. Based on the experience the team had in analyzing litigation data as a measure of quality, the selection of such measures resonates with others such as Wang et al [28], who argued that in the absence of more robust indicators, records of medical malpractice litigation in China warranted further exploration as an indicator of health care quality. Additional measures of clinical quality were accessed by reviewing hospital standards and accreditation of treatment excellence performance against the National Health Commission's Diagnosis and Treatment Improvement of Rare Diseases Program and National Key Clinical Specialty Program.

Developing a Hospital Ranking System

Several rounds of expert consultation identified publicly available indicators, deliberated their utility, and assessed how best to triangulate and weight these measures into comparative performance information. Following the identification and clustering of the available performance measures, we set out to translate these into a practical performance ranking table to assess variation in hospital performance.

Our analysis of operational size and scale highlighted practical limitations to the sample of hospitals included in our ranking. As a result, we focused on the tertiary hospital sector based on the availability of current data as well as to provide an option for comparison with other available measures. By the end of 2017, according to the China Health Statistics Yearbook, there were 1360 grade III, level A hospitals nationwide [27]. The inclusion of these hospitals over others was on the basis that these organizations continue to be the focus of attention in China given their prominence and popularity. These hospitals have also been the focus on other performance rankings in China; hence, the development of any new ranking system would be comparable with other respective performance measures. Our inclusion criteria, therefore, required hospitals to be a grade III, level A hospital, featuring on one of the lists of the four Chinese Hospital rankings in any previous year, and have at least 500 beds. A total of 310 hospitals were thus deemed eligible for ranking under the full criteria.

To develop our ranking system, we relied on statistical procedures such as principal component analysis (PCA) and categorical principal component analysis (CATPCA) [30,31]. PCA is defined as a variable reduction technique that can be used when variables are highly interrelated, providing a way to reduce the number of observed variables into a smaller number of linear, uncorrelated summary variables called principal components (PCs) that account for variation in observed variables. Here, we hypothesized that the various candidate indicators for a given hospital can be represented by several underlying, or latent PCs that reflect the overall strength of this hospital. Thus, for each PC, the model can estimate the extent to which the values are the result of a relationship with the composite score. The remaining variance in the indicators is attributed to measurement error. The degree to which an indicator is correlated with other indicators helps to determine its weight in the equation for the composite scores.

We developed PCA/CATPCA models for each of the three domains of ranking, by evaluating model statistics for all possible combinations of indicators that included at least one indicator. From the resulting list of candidate models showing acceptable fit statistics, we selected a final model for each domain, providing a combination of the number of indicators (models with more indicators produce more accurate component scores), number of outcomes, and model fit.

The PCA/CATPCA replaces the original n features with a smaller number of m features, which are linear combinations of the old features, making these linear combinations as irrelevant as possible. To adjust for linear distribution, before incorporating the values of PCs into the scoring, we implemented a logit transformation for each domain. Equation (1) presents the formula for logit transformation:

Score(i): final score for PC_i; i: PC_i; G(m): accumulation of variance; C_i: variance for PC_i; and F_i: original score for PC_i.

Entropy Weight Method

For each domain, we incorporated the values of PCs into the scoring by using entropy weight method, in which weights are systematically calculated based on the level of the difference between the original values. Simply put, if the value difference between the objects, when evaluated using an indicator, is higher than the difference using other indicators, that indicator has more weight than other indicators [32].

Matrix after logical transformation:

where n: number of hospitals; m: number of variables.

Weighting

Deliberations between the expert panel of HDI stakeholders determined the appropriate weights for each domain based on their importance in defining the overall attributes of strength within hospitals.

For presentation purposes, we created what we define as CHDI to measure the development level of the hospitals evaluated. Raw scores were transformed to a scale that assigns a CHDI score of 100 to the top hospital. The formula for the transformation is shown in Equation (7):

CHDI score = (raw score – minimum) / range (7)

Before applying PCA, we also measured the correlation of each variable using Kaiser–Meyer–Olkin (KMO) analysis. If the KMO value is >.7, there is a relatively high level of correlation among variables, and it is thus suitable to use PCA. Similarly, we calculated Cronbach α coefficient before applying CATPCA. CATPCA is an alternative to standard PCA that is particularly useful for data sets consisting of categorical variables (nominal or ordinal) that might be nonlinearly related to each other. CATPCA quantifies categorical variables using optimal scaling, resulting in optimal PCs for the transformed variables. The correlations, shown in Table 3, provide strong evidence of construct validity.

Table 3. Kaiser–Meyer–Olkin (KMO) analysis (Cronbach alpha values) of the three domains of the China Hospital Development Index.

Domain	KMO/Cronbach α
Research and development	.850
Academic reputation	.802
Quality and safety	.936

Results

PCA Results

Table 4 shows the results of the analysis, including the number of original indicators, number of selected indicators, number of PCs retained, and accumulation of variance for each domain. The PCA resulted in two PCs, which explained no less than 81% of the variance in the original matrix for each domain.

For research and development, the first principal component (PC1) is highly correlated with the amount of SCI papers and citations, whereas the second principal component (PC2) is highly correlated with high-quality paper measures such as “number of IF≥10 SCI papers by authors” and “number of IF≥10 SCI papers by first author” with correlation coefficients of 0.489 and 0.575, respectively, between the original variables and PCs identified. For quality and safety, PC1 is highly correlated with the medical malpractice claims measures, whereas PC2 is highly correlated with the quality of specialty care measures.

Table 4. Principal component analysis and categorical principal component analysis results of the three domains of the China Hospital Development Index.

Domain	Original indicators	Selected indicators	Principal component	Accumulation of variance
Research and development	13	11	2	0.936
Academic reputation	9	6	2	0.818
Quality and safety	3	3	2	0.887

Results of Entropy Weight Method

Table 5 shows PC1 has a higher entropy weight than PC2 for each domain, suggesting that this component has a bigger difference.

The results of our CHDI rankings by score for the top 10 hospitals are shown in Table 6.

Table 5. Results of the entropy weight method for the three domains of the China Hospital Development Index.

Domain	PC1[a] (%)	PC2[b] (%)
Research and development	85.4	14.6
Academic reputation	70.2	29.8
Quality and safety	85.1	14.9

[a] PC1: first principal component.

[b] PC2: second principal component.

Table 6. An example of the China Hospital Development Index (CHDI) ranking results.

Rank	Hospital	CHDI
1	XH Hospital, Beijing	0.9739
2	HX Hospital, Chengdu	0.9635
3	JZ Hospital, Beijing	0.9634
4	RJ Hospital, Shanghai	0.9632
5	ZS Hospital, Shanghai	0.9527
6	HS Hospital, Shanghai	0.9400
7	ZY Hospital, Hangzhou	0.9284
8	CH Hospital, Shanghai	0.9206
9	RM Hospital, Beijing	0.9171
10	BY Hospital, Beijing	0.9137

Validation

Table 7 shows the correlation between scores of domains for 310 hospitals. All of the correlation coefficients between the total score and the score of each domain are above 0.57. The scores of different domains also correlate well among themselves with correlation coefficients higher than 0.50, indicating that the set of indicators is compact and coherent.

Table 7. Correlation coefficients (*r*) between scores of the three domains for all hospitals evaluated (N=310).

Domain	Research and development	Academic reputation	Quality and safety	Total score
Research and development				
r	1	0.788	0.577	0.959
P value	—[a]	<.001	<.001	<.001
Academic reputation				
r	0.788	1	0.597	0.857
P value	<.001	—	<.001	<.001
Quality and safety				
r	0.577	0.597	1	0.773
P value	<.001	<.001	—	<.001
Total score				
r	0.959	0.857	0.773	1
P value	<.001	<.001	<.001	—

[a] Not applicable.

Discussion

Principal Findings

The ability to assess the performance of hospitals in supporting the delivery of high-quality patient care represents a priority for all health care systems [33]. In China, such interest and scrutiny are growing; however, gaining access to relevant performance information remains challenging [34]. Through our analysis of available performance measures, our study aims to contribute to these discussions and debates with a review of hospital performance measures in China. Compared to other health care systems, most notably those in the United States and Europe, what these various measures show are clear limitations in what is currently available to understand hospital performance. For example, our use of research and development indicators and academic reputation as proxy measures for managerial and clinical leadership are exposed to criticisms for their limited connections to day-to-day hospital practice. Our use of litigation data and accreditation standards as proxy measures for hospital quality and safety again have limitations in terms of how far these reflect the clinical quality of hospital care [35]. There is further work to be done regarding how China can develop more clinically focused performance measures that are comparable across hospitals.

Nevertheless, we would argue that our review and subsequent development of a new ranking system breaks new ground in understanding hospital performance in China. Where current hospital rankings in China often rely on reputation scores and investment (input) measures [8,11,12], our review of publicly available measures and their development into a ranking system appears to be the first in opening up the debate for more rigorous and transparent performance information. This is particularly the case for quality and safety performance. Our review and inclusion of litigation data [28,29,36,37] provides a valuable opportunity to assess the comparative performance of quality and safety across hospitals in China.

Thus, this paper contributes to what appears to be a growing body of knowledge that is using innovative and feasible methodologies in data collection and modeling to better understand the performance of public hospitals in China [38].

Based on our analysis, we suggest that further research and policy development is needed to build on these results. Given the practical limitations of securing comparative data and the interest in benchmarking our analysis with existing hospital rankings in China, our sample has focused exclusively on a number of tertiary hospitals. Our rankings reflect the high performance of these organizations compared to that of other hospital and primary care providers; however, we are also mindful of the possible further imbalance this can create in China's health care system by virtue of acute medical care over primary and community care. The fact that the majority of our highest ranked hospitals are located in Shanghai and Beijing also illustrates important challenges facing access to high-quality hospital care in other parts of China. Such findings support those of Yu et al [38] who have documented how the unevenness of health care resources in China is closely related to a city's administrative rank and power: the higher the level, the better the resources. Such arrangements are reinforcing investment in high-ranked hospitals at the expense of primary care services. The correlation between the quality and safety domain and the overall hospital performance in our ranking system is slightly lower than that with the other two domains; therefore, more clinical objective measures should be included to increase the influence of this domain.

Therefore, we call on further research and development to access and compare performance measures from within each hospital, including private hospitals, as well as other parts of the health care system, including primary and community care. For this purpose, China could build on the cross-sectional research it has undertaken into mortality trends [39] and nurse staffing levels [40]. Such research has the potential to be scaled up and developed into performance measures translatable across all hospitals and incorporated into our methodology.

We also support further research and development that draws on the views of a range of different stakeholders in terms of what performance measures would be meaningful for patients, public, and health care staff. Given the well-documented challenges facing the doctor-patient relationship in China [41,42], we encourage deliberative events involving a range of stakeholders to discuss what constitutes good performance with the view to developing shared understanding of performance measurement from different perspectives. The Delphi method [43] is one way to do this, and such an approach has been used in other parts of China to good effect. This includes further development of comparative measures for health outcomes and the development of experiential data about how different stakeholders experience the health care received and how they can improve hospital performance as well as other aspects of China's health care system.

The year 2019 marks the tenth year for China's goal to deepen the reform of its medical and health system. In 2019, the Chinese government formulated the National Tertiary Public Hospital Performance evaluation index system and unified the collection of performance information across hospitals [44]. The implications of such changes remain to be seen, with the results of these assessments not yet fully disclosed. However, we anticipate this is an important step in developing greater understanding of hospital performance in China. We believe that our review and the newly developed ranking index (CHDI) has an important role to play in shaping such discussions and assessments, particularly in relation to the improvement of quality and patient safety, as well as raising public awareness regarding the information that is available to inform their decision-making.

Conclusions

The reform of China's health care system has brought with it increasing scrutiny regarding the quality of care delivered to patients. Our analysis presents what appears to be the first review of publicly available performance measures for hospitals in China. In collaboration with an expert panel, in the review, the available measures have been clustered into three performance domains, namely research and development, academic reputation, and quality and safety of hospital care. Furthermore, based our analysis, we have applied these performance measures to a selection of tertiary hospitals in China with the view to better understand their comparative performance. There remain some notable limitations and challenges regarding this performance information; nevertheless, we believe that our review and ranking system break new ground in assessing hospital performance in China. Although further research and development is clearly needed to enhance and refine this performance information, we argue that the proposed hospital development index sets a new and important research agenda for understanding and improving hospital care in China.

Acknowledgments

The authors acknowledge the research institution, university, and the data collectors for their participation and support in this study. This study was supported by Major Key Research Projects in Philosophy and Social Sciences of the Ministry of Education of the People's Republic of China (grant 18JZD044) and National Natural Science Foundation of China (grant 72074147).

Authors' Contributions

SJD participated in the acquisition of data, design and analysis of the study, and drafted the manuscript. RM was involved in drafting the final manuscript. CSS contributed to the conception and design of the study. MYD and YYX assisted with data collection and statistical analysis of the data. GHL, JS, and XQF prompted discussion on the implications of the study and its results and critically revised the manuscript for important intellectual content. All authors have read and approved the final version of the manuscript and are accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ranking process of China Hospital Development Index.
[DOCX File , 34 KB - jmir_v23i6e17095_app1.docx]

References

1. Leonardi MJ, McGory ML, Ko CY. Publicly available hospital comparison web sites: determination of useful, valid, and appropriate information for comparing surgical quality. Arch Surg 2007 Sep;142(9):863-8; discussion 868. [doi: 10.1001/archsurg.142.9.863] [Medline: 17875841]
2. Hwang W, Derk J, LaClair M, Paz H. Finding order in chaos: a review of hospital ratings. Am J Med Qual 2016;31(2):147-155. [doi: 10.1177/1062860614558085] [Medline: 25381001]
3. Huerta TR, Walker DM, Ford EW. An evaluation and ranking of children's hospital websites in the United States. J Med Internet Res 2016 Aug 22;18(8):e228 [FREE Full text] [doi: 10.2196/jmir.5799] [Medline: 27549074]
4. Bilimoria K, Birkmeyer J, Burstin H. Rating the raters: an evaluation of publicly reported hospital quality rating systems. NEJM Catalyst 2019 Aug 14.
5. Mei J, Kirkpatrick I. Public hospital reforms in China: towards a model of new public management? IJPSM 2019 May 13;32(4):352-366. [doi: 10.1108/ijpsm-03-2018-0063]
6. Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 years of health-care reform in China: progress and gaps in Universal Health Coverage. Lancet 2019 Sep 28;394(10204):1192-1204. [doi: 10.1016/S0140-6736(19)32136-1] [Medline: 31571602]
7. Pan J, Qin X, Hsieh C. Is the pro-competition policy an effective solution for China's public hospital reform? HEPL 2016 Jun 27;11(4):337-357. [doi: 10.1017/s1744133116000220]
8. Wu S, Chen T, Pan Q, Wei L, Xuan Y, Li C, et al. Establishment of a comprehensive evaluation system on medical quality based on cross-examination of departments within a hospital. Chin Med J (Engl) 2017 Dec 05;130(23):2872-2877 [FREE Full text] [doi: 10.4103/0366-6999.219163] [Medline: 29176146]
9. FAQ: how and why we rank and rate hospitals. U.S. News. 2020 Jul 28. URL: https://health.usnews.com/health-care/best-hospitals/articles/faq-how-and-why-we-rank-and-rate-hospitals [accessed 2020-07-30]
10. Member awards: a different approach for 2020. Vizient. URL: https://www.vizientinc.com/members/member-awards [accessed 2020-07-30]
11. Xiong X, Gao J, Zhang B, Lu J, Luo L. Introduction of Chinese hospital ranking method from the aspect of theoretical framework, practical choice and social effect. J Hosp Manag Health Policy 2017 Nov 6;1:4-400. [doi: 10.21037/jhmhp.2017.11.01]
12. Dai T, Qian Q, Xin-Ying AN. Research on the Sci-tech Influence Evaluation Index System for hospitals in China. Article in Chinese. Journal of Medical Informatics 2016;37(3):2-7 [FREE Full text]
13. Chong E, Yan A, Wang S. Research on the comprehensive ranking method of hospital competitiveness. Article in Chinese. Chinese Journal of Medical Management Sciences 2016;6(3):45-50 [FREE Full text]
14. Peking University released "China's Best Clinical Discipline Rankings". People's Daily Online. 2015 Jun 4. URL: http://scitech.people.com.cn/n/2015/0604/c1057-27100408.html [accessed 2021-05-31]
15. Jian W, Figueroa J, Woskie L, Yao X, Zhou Y, Li Z, et al. Quality of care in large Chinese hospitals: an observational study. BMJ Qual Saf 2019 Dec;28(12):963-970. [doi: 10.1136/bmjqs-2018-008938] [Medline: 31110140]
16. Wang DE, Tsugawa Y, Figueroa JF, Jha AK. Association between the Centers for Medicare and Medicaid Services hospital star rating and patient outcomes. JAMA Intern Med 2016 Jun 01;176(6):848-850. [doi: 10.1001/jamainternmed.2016.0784] [Medline: 27062275]
17. Web of Science. Clarivate. URL: https://clarivate.com/webofsciencegroup/solutions/web-of-science/ [accessed 2021-05-26]
18. Chinese Clinical Trial Registry. URL: http://www.chictr.org.cn/enIndex.aspx [accessed 2021-05-26]
19. Chinese Academy of Sciences. Webpage in Chinese. URL: https://www.cas.cn/ [accessed 2021-05-26]
20. Chinese Academy of Engineering. URL: http://en.cae.cn/en/ [accessed 2021-05-26]
21. Chinese Science Citation Database, Science China. URL: http://www.sciencechina.cn/index.jsp [accessed 2021-05-26]
22. Chinese Medical Association. URL: http://en.cma.org.cn/ [accessed 2021-05-26]
23. Chinese Medical Doctor Association. URL: http://www.cmdae.org/ [accessed 2021-05-26]

24. National Office for Science and Technology Awards. URL: http://www.nosta.gov.cn/english/index.html [accessed 2021-05-26]
25. National Health Commission of the People's Republic of China. URL: http://en.nhc.gov.cn/ [accessed 2021-05-26]
26. Laws and Regulations – Peking University. Webpage in Chinese. URL: www.pkulaw.cn [accessed 2021-05-26]
27. China Statistical Yearbook 2018. Beijing: China Statistics Press; 2018.
28. Wang Z, Li N, Jiang M, Dear K, Hsieh C. Records of medical malpractice litigation: a potential indicator of health-care quality in China. Bull World Health Organ 2017 Jun 01;95(6):430-436 [FREE Full text] [doi: 10.2471/BLT.16.179143] [Medline: 28603309]
29. Järvelin J, Häkkinen U. Can patient injury claims be utilised as a quality indicator? Health Policy 2012 Feb;104(2):155-162. [doi: 10.1016/j.healthpol.2011.08.012] [Medline: 21956047]
30. Linting M, van der Kooij A. Nonlinear principal components analysis with CATPCA: a tutorial. J Pers Assess 2012;94(1):12-25. [doi: 10.1080/00223891.2011.627965] [Medline: 22176263]
31. Meulman J, Heiser W. Categorical principal component analysis (CATPCA). SPSS Categories 2005:14.
32. Jing N, Wang J. Evaluate the International Talents in Professional Social Networks using the Entropy Weight MethodC//2015 International Conference on Computational Science and Computational Intelligence (CSCI). 2015 Presented at: International Conference on Computational Science and Computational Intelligence; 2015; Las Vegas, NV, USA p. 220-225. [doi: 10.1109/csci.2015.113]
33. Li X, Krumholz HM. What does it take to improve nationwide healthcare quality in China? BMJ Qual Saf 2019 Dec;28(12):955-958. [doi: 10.1136/bmjqs-2019-009839] [Medline: 31366577]
34. Li C, Xu X, Zhou G, He K, Qi T, Zhang W, et al. Implementation of National Health Informatization in China: survey about the status quo. JMIR Med Inform 2019 Feb 21;7(1):e12238 [FREE Full text] [doi: 10.2196/12238] [Medline: 30789350]
35. Garon-Sayegh P. Analysis of medical malpractice claims to improve quality of care: Cautionary remarks. J Eval Clin Pract 2019 Oct;25(5):744-750. [doi: 10.1111/jep.13178] [Medline: 31069900]
36. Li H, Wu X, Sun T, Li L, Zhao X, Liu X, et al. Claims, liabilities, injures and compensation payments of medical malpractice litigation cases in China from 1998 to 2011. BMC Health Serv Res 2014 Sep 13;14:390 [FREE Full text] [doi: 10.1186/1472-6963-14-390] [Medline: 25218509]
37. Greenberg MD, Haviland AM, Ashwood JS, Main R. Is better patient safety associated with less malpractice activity?: Evidence from California. Rand Health Q 2011;1(1):1 [FREE Full text] [Medline: 28083157]
38. Yu M, He S, Wu D, Zhu H, Webster C. Examining the multi-scalar unevenness of high-quality healthcare resources distribution in China. Int J Environ Res Public Health 2019 Aug 07;16(16):2813 [FREE Full text] [doi: 10.3390/ijerph16162813] [Medline: 31394765]
39. Zhang Z, Lei J, Shao X, Dong F, Wang J, Wang D, China Venous Thromboembolism Study Group. Trends in hospitalization and in-hospital mortality from VTE, 2007 to 2016, in China. Chest 2019 Feb;155(2):342-353. [doi: 10.1016/j.chest.2018.10.040] [Medline: 30419233]
40. Zhu X, You L, Zheng J, Liu K, Fang J, Hou S, et al. Nurse staffing levels make a difference on patient outcomes: a multisite study in Chinese hospitals. J Nurs Scholarsh 2012 Sep;44(3):266-273. [doi: 10.1111/j.1547-5069.2012.01454.x] [Medline: 22732012]
41. He AJ, Qian J. Explaining medical disputes in Chinese public hospitals: the doctor-patient relationship and its implications for health policy reforms. Health Econ Policy Law 2016 Oct;11(4):359-378. [doi: 10.1017/S1744133116000128] [Medline: 27018911]
42. Hesketh T, Wu D, Mao L, Ma N. Violence against doctors in China. BMJ 2012 Sep 07;345:e5730. [doi: 10.1136/bmj.e5730] [Medline: 22960376]
43. Hu Q, Qin Z, Zhan M, Wu B, Chen Z, Xu T. Development of a trigger tool for the detection of adverse drug events in Chinese geriatric inpatients using the Delphi method. Int J Clin Pharm 2019 Oct;41(5):1174-1183. [doi: 10.1007/s11096-019-00871-x] [Medline: 31254152]
44. Opinions of the General Office of the State Council on strengthening the performance appraisal work of tertiary public hospitals. Webpage in Chinese. Office of the State Council, People's Republic of China. 2019 Jan 30. URL: http://www.gov.cn/zhengce/content/2019-01/30/content_5362266.htm [accessed 2021-05-25]

Abbreviations

CAE: Chinese Academy of Engineering
CAS: Chinese Academy of Sciences
CATPCA: categorical principal component analysis
CDA: Chinese Doctor Award
CHDI: China Hospital Development Index
CMA: Chinese Medical Association
CMDA: Chinese Medical Doctor Association
CMSTA: Chinese Medical Science and Technology Award

CSCD: China Science Citation Database
HDI: Hospital Development Institution
IF: impact factor
KMO: Kaiser–Meyer–Olkin
PC: principal component
PC1: first principal component
PC2: second principal component
PCA: principal component analysis
SCI: Science Citation Index
SNSA: State Natural Science Award
SPSTA: State Preeminent Science and Technology Award
STEM: Science and Technology Evaluation Metrics
STTPA: State Scientific and Technological Progress Award

Edited by G Eysenbach, R Kukafka; submitted 19.11.19; peer-reviewed by W Zhang, R Waitman; comments to author 29.06.20; revised version received 12.08.20; accepted 30.04.21; published 17.06.21.

Please cite as:
Dong S, Millar R, Shi C, Dong M, Xiao Y, Shen J, Li G
Rating Hospital Performance in China: Review of Publicly Available Measures and Development of a Ranking System
J Med Internet Res 2021;23(6):e17095
URL: https://www.jmir.org/2021/6/e17095
doi:10.2196/17095
PMID:34137724

Original Paper

The Association Between Logging Steps Using a Website, App, or Fitbit and Engaging With the 10,000 Steps Physical Activity Program: Observational Study

Anna T Rayward[1,2], PhD; Corneel Vandelanotte[1], PhD; Anetta Van Itallie[1], BHMS; Mitch J Duncan[3], PhD

[1]School of Health, Medical and Applied Sciences, Central Queensland University, Rockhampton, Australia
[2]School of Education, College of Human and Social Futures, University of Newcastle, Callaghan, Australia
[3]School of Medicine & Public Health, College of Health, Medicine and Wellbeing, University of Newcastle, Callaghan, Australia

Corresponding Author:
Anna T Rayward, PhD
School of Health
Medical and Applied Sciences
Central Queensland University
Bruce Highway
Rockhampton, 4700
Australia
Phone: 61 240553239
Email: anna.rayward@newcastle.edu.au

Abstract

Background: Engagement is positively associated with the effectiveness of digital health interventions. It is unclear whether tracking devices that automatically synchronize data (eg, Fitbit) produce different engagement levels compared with manually entering data.

Objective: This study examines how different step logging methods in the freely available 10,000 Steps physical activity program differ according to age and gender and are associated with program engagement.

Methods: A subsample of users (n=22,142) of the free 10,000 Steps physical activity program were classified into one of the following user groups based on the step-logging method: Website Only (14,617/22,142, 66.01%), App Only (2100/22,142, 9.48%), Fitbit Only (1705/22,142, 7.7%), Web and App (2057/22,142, 9.29%), and Fitbit Combination (combination of web, app, and Fitbit; 1663/22,142, 7.51%). Generalized linear regression and binary logistic regression were used to examine differences between user groups' engagement and participation parameters. The time to nonusage attrition was assessed using Cox proportional hazards regression.

Results: App Only users were significantly younger and Fitbit user groups had higher proportions of women compared with other groups. The following outcomes were significant and relative to the Website Only group. The App Only group had fewer website sessions (odds ratio [OR] −6.9, 95% CI −7.6 to −6.2), whereas the Fitbit Only (OR 10.6, 95% CI 8.8-12.3), Web and App (OR 1.5, 95% CI 0.4-2.6), and Fitbit Combination (OR 8.0; 95% CI 6.2-9.7) groups had more sessions. The App Only (OR −0.7, 95% CI −0.9 to −0.4) and Fitbit Only (OR −0.5, 95% CI −0.7 to −0.2) groups spent fewer minutes on the website per session, whereas the Fitbit Combination group (OR 0.2, 95% CI 0.0-0.5) spent more minutes. All groups, except the Fitbit Combination group, viewed fewer website pages per session. The mean daily step count was lower for the App Only (OR −201.9, 95% CI −387.7 to −116.0) and Fitbit Only (OR −492.9, 95% CI −679.9 to −305.8) groups but higher for the Web and App group (OR 258.0, 95% CI 76.9-439.2). The Fitbit Only (OR 5.0, 95% CI 3.4-6.6), Web and App (OR 7.2, 95% CI 5.9-8.6), and Fitbit Combination (OR 15.6, 95% CI 13.7-17.5) groups logged a greater number of step entries. The App Only group was less likely (OR 0.65, 95% CI 0.46-0.94) and other groups were more likely to participate in Challenges. The mean time to nonusage attrition was 35 (SD 26) days and was lower than average in the Website Only and App Only groups and higher than average in the Web and App and Fitbit Combination groups.

Conclusions: Using a Fitbit in combination with the 10,000 Steps app or website enhanced engagement with a real-world physical activity program. Integrating tracking devices that synchronize data automatically into real-world physical activity interventions is one strategy for improving engagement.

(J Med Internet Res 2021;23(6):e22151) doi:10.2196/22151

KEYWORDS

physical activity intervention; activity trackers; engagement; Fitbit; pedometer; eHealth; mobile phone

Introduction

Background

Meeting or exceeding recommended physical activity levels is key for the prevention and management of noncommunicable diseases [1,2]. However, large proportions of the population do not meet these recommendations [3,4]. In response, web- and app-based programs to promote physical activity among adults have been developed [5]. Few of these interventions have been evaluated in real-world settings [6]. Relative to randomized controlled trials whose participants are rigorously screened, have repeated contact with trial staff, and may include participants who are motivated to change behaviors, interventions conducted in less controlled and real-world settings may have significantly different levels of usage, engagement, nonusage attrition (users stop interacting entirely), and behavior change [6]. However, the greater accessibility of real-world programs makes them valuable avenues to reach larger populations at a relatively low cost [7]. Furthermore, the way users engage with web- and app-based programs in ecologically valid circumstances may potentially impact the effectiveness of these programs [6].

Web- and app-based programs that promote physical activity frequently include self-monitoring as an effective behavior change technique [8]. Although pedometers are commonly and successfully used in activity promotion efforts [9], the emergence of wrist-worn activity trackers such as Fitbit has provided a convenient, reliable, and accurate alternative to pedometers for tracking step counts. Worldwide sales of health and fitness trackers have more than tripled from 2014 (26 million units) to 2017 (87 million units) [10]. The use of technology to track health is widespread, with 33% of the worldwide population across all age groups using a mobile app or a fitness tracking device in 2016 [11]. Australia had the second highest wearable fitness device adoption rate in the world in 2016, with 14% of the population owning at least 1 device [12]. Therefore, it is logical to integrate advanced activity trackers that can automatically synchronize activity behavior into web- and app-based programs [13].

A randomized controlled trial, the *TaylorActive* intervention, delivered the same web-based computer-tailored physical activity intervention to a Fitbit group and a non-Fitbit group [14]. Compared with the non-Fitbit group (which self-reported frequency of, and time spent in, mild, moderate, and strenuous physical activity on the website or app), the group using Fitbit activity trackers significantly increased in total weekly physical activity (mean total physical activity increase of 163.2 min/week; 95% CI 52.0-274.0) and weekly moderate-to-vigorous physical activity (mean moderate-to-vigorous physical activity increase of 78.6 min/week; 95% CI 24.4-131.9) after 3 months [14]. However, it is unclear how the addition of a physical activity tracking device impacts the use of free real-world physical activity promotion programs. Previous research has questioned whether there is a difference in the effectiveness of active (eg, manual entry of step count into a database) versus passive (eg, automatic entry of step count into a database by an activity tracking device requiring no cognitive attention) self-monitoring of behaviors. However, there is a paucity of research on this subject [15]. The lack of effort required to log in to a website or app to manually enter recorded physical activity may limit both the cognitive effort and subsequent focus on improving or maintaining an activity goal. It may also preclude opportunities to engage with other program features found on websites and apps, such as goal setting and social support features, which may consolidate behavior change [16,17]. Alternatively, the reduced burden of having to manually log steps may remove perceived barriers, such as time constraints, and subsequently provide more time to engage with program features and result in longer engagement before nonusage attrition occurs. Furthermore, having multiple options on how to self-monitor physical activity could overwhelm some users, but it may increase interest and better engagement in other users through enhanced accessibility of their preferred self-monitoring method. In addition, fitness trackers are used more commonly by women and younger adults, which may influence engagement with web-based health interventions [10,11,18].

The 10,000 Steps program is a free, publicly available program that aims to promote physical activity through the use of pedometers, activity trackers, a website, and an app [19]. The website and app have been available for public use since 2004 and 2012, respectively. A previous study examining user engagement with 10,000 Steps program found that program engagement was higher and nonusage attrition was lower among those who used the app or a combination of app and website compared with those using only the website [18]. However, the effect on engagement since the capacity to automatically record and sync step counts with the 10,000 Steps website and app using a Fitbit activity tracking device was introduced in January 2017 has not been assessed.

Objectives

This study aims to examine whether users' logging methods (Website Only, App Only, Fitbit Only, Web and App, and Fitbit Combination) differ according to age and gender as well as how different methods of logging steps in the 10,000 Steps physical activity program are associated with engagement with the program.

Methods

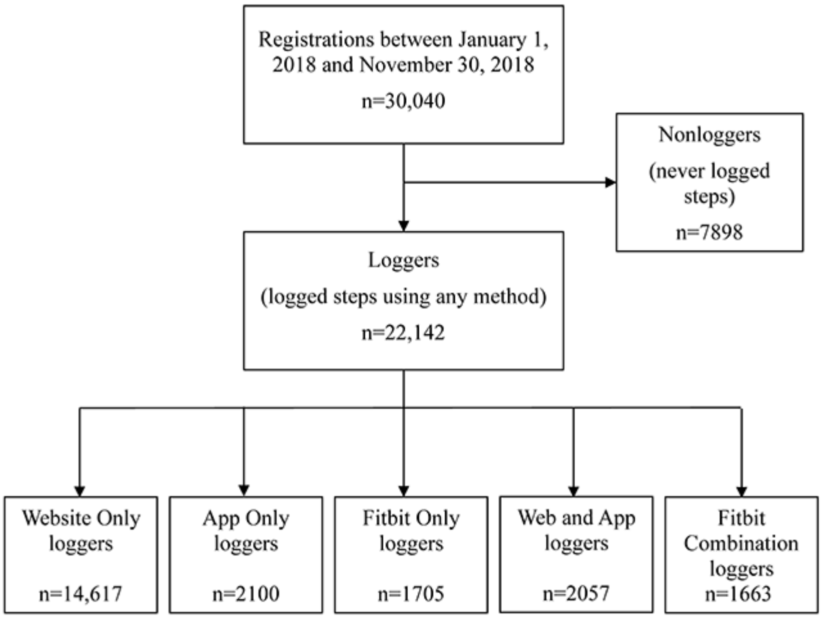
Participants and Procedures

The 10,000 Steps program was designed as a free, publicly available, whole-of-community program to increase physical activity among adults. The program, which is based on a socioecological framework, began in Rockhampton, Australia, and utilizes multiple strategies to promote physical activity [20].

Further details on the program design and development are available elsewhere [21,22]. In general, the program encourages users to accumulate physical activity each day and monitor their daily physical activity levels (actively or passively) by recording their pedometer, activity tracker–counted steps and/or time spent in physical activity using a web-based step log. The web-based step log is available to users on both the 10,000 Steps website and a smartphone app. Activity and/or steps logged using the app are automatically synchronized with the user's account on the website [18]. Steps recorded via a Fitbit activity tracking device were also synchronized with the app and website. Users can also access additional program features, including monthly Challenges for individuals (users may choose from a selection of virtual journeys with predefined monthly step goals and receive feedback in relation to progress), team Tournaments (created by 10,000 Steps coordinators from workplaces, community organizations, or groups of friends that involve team-based virtual walking Challenges based on a set amount of time or a predefined journey), and virtual *friends* (which allow users to track one another's progress and motivate each other) [18].

In this study, participants were users of the 10,000 Steps program who registered between January 1, 2018, and November 30, 2018 (N=30,040). When registering with the 10,000 Steps program, users provided informed consent for the usage of their data for research purposes. Of these new registrations, 7898 never logged steps (referred to as *Nonloggers*) and were not included in the analyzed sample. The remaining 22,142 new users were classified into one of five user groups based on the method they used to log steps. Participants captured their daily step counts using either a Fitbit or some other device (eg, pedometers, phone apps, or non-Fitbit activity monitors). Then, they recorded their step counts on the 10,000 Steps platform by manually entering data via the website or app or via automatic syncing of their Fitbit with the platform. The methods used to log step entries were recorded on a website database. The user groups were as follows: Website Only (14,617/22,142, 66.01%; step count data entered *only via* the website), App Only (2100/22,142, 9.48%; step count data entered *only* via the app), Fitbit Only (1705/22,142, 7.7%; step count data entered *only* via the Fitbit), Web and App (2057/22,142, 9.29%; entered step count data using the website and the app), and Fitbit Combination (1663/22,142, 7.51%; step count data entered using a combination of Fitbit and website and/or app). The Fitbit Only group logged steps passively (although may have actively engaged with the website content otherwise) while the Fitbit Combination group logged steps passively and actively. All other groups actively logged the steps. Regardless of how users logged steps, they were all able to interact with the website content (Figure 1).

Figure 1. Classification of logging methods of new registrations to the 10,000 Steps program between January 1, 2018, and November 30, 2018.



Data Collection and Extraction

Data were extracted from the 10,000 Steps website database and Google Analytics to assess user characteristics, engagement, and platform usage. Data for new users were tracked from the time of registration (between January 1 and November 30, 2018) until December 31, 2018, which allowed a minimum of 31 days opportunity to observe all participants' self-monitoring of physical activity.

Measures

User Characteristics

The date of birth, gender, and registration date were recorded and stored in the website database. The date of birth was used to determine the age at the time of registration. Gender was categorized as male, female, or other.

User Engagement With Website and Program

User engagement data were extracted for the entire study period from January 1 to December 31, 2018. User groups were mutually exclusive and were defined based on the method of logging steps. However, irrespective of a user's logging method, all users had the potential to access the website (eg, Fitbit Only users only logged steps via a Fitbit and had the opportunity to view the 10,000 Steps website and/or app). The website includes a variety of information, including library articles (eg, benefits of physical activity and strategies for increasing activity) with information on workplace and individual Challenges [22]. Users' engagement with the 10,000 Steps website was assessed based on all pages or sections of the website by using the average number of sessions (using the website to interact with content; eg, viewing step statistics and joining a Challenge), average time per session, number of pages viewed, average number of pages viewed per session, average number of step log entries, and average daily step count on days when steps were logged. Participating in individual Challenges and team Tournaments and receiving or sending friend requests were all dichotomized as *yes* or *no* (yes=participated in ≥1 Challenge or Tournament; yes=received or sent ≥1 friend request). Any average daily step count of more than 20,000 steps was truncated to 20,000 steps (n=1147) [18]. Nonusage attrition was classified as having no data entries for ≥14 consecutive days. The time to nonusage attrition was calculated using data exported from the website [23].

Statistical Analysis

Results were expressed as differences in least square group means (odds ratios [ORs]) or hazard ratios (HRs) with 95% CIs. α was set at .05 for all analyses, which were conducted using Stata V15.1 (StataCorp LLC). in August 2019 [24]. Summary statistics were used to describe user characteristics. Group comparisons regarding age were assessed using 1-way analysis of variance with Bonferroni correction. Group comparisons regarding gender and how users found the program were assessed using chi-square tests.

Generalized linear models were used to examine differences between user groups in website sessions, minutes per session, page views, pages per session, daily step count, and total step log entries. Negative binomial family and identity link functions were used for all analyses except step count, which used a Poisson family and identity link model. The specification of the family and link functions was informed by residual diagnostics. All models were adjusted for participation in Challenges and Tournaments, as these features have been shown to be associated with usage in previous studies of 10,000 Steps users [18,25].

Associations between user groups in participation in Challenges and Tournaments and receiving and sending friend requests were assessed using binary logistic regression. These models were adjusted for Challenges and/or Tournaments depending on the outcome examined (eg, when Challenges were the outcome, model adjusted for participation in Tournament; when Tournaments were the outcome, model adjusted for participation in Challenges).

Usage and engagement with the 10,000 Steps program have been shown to be affected by participation in Challenges and Tournaments [18,25]. Consequently, 2 Cox proportional hazards regressions were conducted to assess between-group differences in time to nonusage attrition, first using an unadjusted model and then using a model adjusting for participation in Challenges and Tournaments. User group time to nonusage attrition was plotted using Kaplan-Meier survival estimates and adjusted survival curves. No obvious violations were observed upon examination of Schoenfeld residuals.

Results

Overview

Of the 30,040 participants', whose mean age was 40 years, 20,992 (69.88%) were women and 7898 never logged steps (Table 1).

Table 1. Descriptive summary of new users—by user group—of the 10,000 Steps program registered between January 1 and November 30, 2018 (N=30,040).

Characteristics	Website Only (n=14,617)	App Only (n=2100)	Fitbit Only (n=1705)	Web and App (n=2057)	Fitbit Combination (n=1663)	Nonlogger[a] (n=7898)
Age (years)[b,c], mean (SD)	40.45 (12.5)	37.7 (11.8)	39.9 (11.3)	38.9 (11.6)	40.5 (12.0)	41.0 (14.0)
Gender[d], n (%)						
Female	9837 (67.3)	1363 (65)	1279 (75)	1391 (67.6)	1285 (77.3)	5837 (73.9)
Male	4620 (31.6)	725 (34.5)	409 (24)	653 (31.8)	367 (22.1)	2015 (25.5)
Other	160 (1.1)	12 (0.6)	17 (1)	13 (0.6)	11 (0.7)	46 (0.6)
Ever had a website session, n (%)	14,355 (98.2)	1938 (92.3)	1696 (99.5)	2028 (99.5)	1645 (98.6)	4046 (51.2)
Number of website sessions, mean (SD)	18.1 (22.8)	10.9 (20.4)	33.5 (46.3)	21.7 (27.4)	31.5 (40.5)	1.8 (2.0)
Session duration (min/session), mean (SD)	5.9 (4.9)	5.2 (6.0)	5.4 (4.5)	5.8 (4.6)	6.0 (4.4)	7.2 (8.7)
Number of page views, mean (SD)	130.7 (144.3)	50.6 (73.3)	199.5 (233.0)	133.6 (160.6)	213.0 (240.0)	16.3 (21.5)
Number of pages viewed per session, mean (SD)	9.1 (5.5)	6.7 (4.3)	7.9 (5.5)	7.9 (4.6)	8.6 (4.7)	9.9 (8.1)
Number of step log entries, mean (SD)	31.45 (24.7)	32.7 (25.1)	40.6 (36.4)	41.0 (33.6)	54.9 (49.4)	0 (0.0)
Mean daily steps[e], mean (SD)	10,957 (4172)	10,773 (4038)	10,553 (3699)	11,307 (3887)	11,145 (3516)	0 (0.0)
Participated in at least 1 Challenge, n (%)	353 (2.4)	33 (1.6)	125 (7.3)	110 (5.3)	157 (9.4)	1 (<0.1)
Participated in at least 1 Tournament, n (%)	10,740 (73.5)	1635 (77.9)	1323 (77.6)	1691 (82.2)	1370 (82.4)	12 (0.15)
Sent or received in at least one friend request, n (%)	1956 (13.4)	247 (11.8)	363 (21.3)	374 (18.2)	334 (20.1)	1 (0.01)

[a] Some users engaged with program content but never logged steps.

[b] For the *Age* row, n=29,961 because of missing values.

[c] Indicates a significant difference (F_5=26.84; *P*<.001) between usage groups with respect to age.

[d] Indicates a significant difference (X^2_{10}=216.7; *P*<.001) between usage groups with respect to gender.

[e] The mean daily steps include steps allocated as a result of additionally recorded moderate or vigorous activity and/or distance.

User Characteristics

Table 1 shows a descriptive summary of user groups related to age and gender. App Only users were significantly younger than all other groups (Website Only: *P*<.001; Fitbit Only: *P*<.001; Fitbit Combination: *P*<.001; Nonloggers: *P*<.001; Web and App users: *P*=.05). Web and App users were significantly younger than Website Only users (*P*<.001), Fitbit Combination users (*P*=.001), and Nonloggers (*P*<.001). Nonloggers were significantly older than all other groups except Fitbit Combination users (Website Only: *P*=.01; App Only: *P*<.001; Fitbit Only: *P*=.01; Web and App: *P*<.001). Although statistically significant differences in age were found, they are unlikely to be meaningful because the difference in years was small. Gender was significantly different between user groups with the highest proportions of women in the 2 groups that used Fitbits (Fitbit Only: 1279/1705, 75.01%; Fitbit Combination: 1285/1663, 77.27%).

Between-Group Differences in Engagement With the 10,000 Steps Program

The between-group differences in program engagement metrics from January 1 to December 31, 2018, are shown in Table 2. On average, the App Only group had significantly fewer website sessions (mean difference [MD] −6.9; 95% CI −7.6 to −6.2), whereas the Fitbit Only (MD 10.6; 95% CI 8.8-12.3), Web and App (MD 1.5; 95% CI 0.4- 2.6), and Fitbit Combination (MD 8.0; 95% CI 6.2- 9.7) groups had significantly more website sessions, relative to the Website Only group. The App Only and Fitbit Only groups spent significantly fewer minutes (MD −0.7;

95% CI −0.9 to −0.4; MD −0.5; 95% CI −0.7 to −0.2, respectively), whereas the Fitbit Combination group spent significantly more minutes (MD 0.2; 95% CI 0.0-0.5) on the website for each session, relative to the Website Only group. These differences, although statistically significant, were small in magnitude (<1 min per session).

Table 2. Marginalized means and between–user group differences for website usage and mean daily step count and logging between January 1 and December 31, 2018.

Website usage parameters, step counts, and step log entries[a,b] by group	Value, mean (95% CI)	Between-group coefficient (95% CI)	*P* value[c]
Total number of sessions[d,e]			
Website Only	18.8 (18.4 to 19.1)	Reference category	N/A[f]
App Only	11.9 (11.2 to 12.6)	−6.9 (−7.6 to −6.2)	<.001
Fitbit Only	29.4 (27.7 to 31.1)	10.6 (8.8 to 12.3)	<.001
Web and App	20.2 (19.2 to 21.3)	1.5 (0.4 to 2.6)	.01
Fitbit Combination	26.8 (25.1 to 28.5)	8.0 (6.2 to 9.7)	<.001
Minutes per session[d,e]			
Website Only	5.9 (5.8 to 6.0)	Reference category	N/A
App Only	5.2 (5.0 to 5.5)	−0.7 (−0.9 to −0.4)	<.001
Fitbit Only	5.4 (5.2 to 5.6)	−0.5 (−0.7 to −0.2)	<.001
Web and App	5.9 (5.7 to 6.1)	0.1 (−0.1 to 0.3)	.54
Fitbit Combination	6.1 (5.9 to 6.3)	0.2 (0.0 to 0.5)	.03
Total number of page views[e]			
Website Only	132.3 (130.2 to 134.3)	Reference category	N/A
App Only	62.8 (59.7 to 65.8)	−69.5 (−72.9 to −66.1)	<.001
Fitbit Only	179.9 (171.3 to 188.5)	47.6 (38.8 to 56.5)	<.001
Web and App	123.6 (118.4 to 128.8)	−8.6 (-14.2 to −3.1)	.002
Fitbit Combination	183.0 (175.1 to 191.0)	50.8 (42.5 to 59.0)	<.001
Number of pages viewed per session[d,e]			
Website Only	9.0 (8.9 to 9.1)	Reference category	N/A
App Only	6.8 (6.6 to 7.0)	−2.2 (−2.4 to −2.0)	<.001
Fitbit Only	7.9 (7.6 to 8.1)	−1.2 (−1.4 to −0.9)	<.001
Web and App	8.0 (7.8 to 8.2)	−1.0 (−1.2 to −0.8)	<.001
Fitbit Combination	8.9 (8.7 to 9.1)	−0.1 (−0.4 to 0.1)	.24
Mean daily step count [g]			
Website Only	10,987 (10,920 to 11,055)	Reference category	N/A
App Only	10,785 (10,612 to 10,958)	−201.9 (−387.7 to −116.0)	.03
Fitbit Only	10,494 (10,320 to 10,669)	−492.9 (−679.9 to −305.8)	<.001
Web and App	11,245 (11,077 to 11,413)	258.0 (76.9 to 439.2)	.005
Fitbit Combination	10,996 (10,826 to 11,166)	9.1 (−174.1 to 192.3)	.92
Number of step log entries[e]			
Website Only	32.3 (31.9 to 32.7)	Reference category	N/A
App Only	33.5 (32.3 to 34.6)	1.2 (−0.1 to 2.4)	.07
Fitbit Only	37.3 (35.8 to 38.8)	5.0 (3.4 to 6.6)	<.001
Web and App	39.5 (38.2 to 40.8)	7.2 (5.9 to 8.6)	<.001
Fitbit Combination	47.9 (46.1 to 49.7)	15.6 (13.7 to 17.5)	<.001

[a] Between user groups; Website Only group: n=14,617; App Only group: n=2100; Fitbit Only group: n=1705; Web and App group: n=2057; Fitbit and web and/or app group: n=1663.

[b] Website usage: website usage engagement may not have occurred on consecutive days.

[c] α=.05.

[d] Session: using the website to interact with content. For example, viewing step statistics and joining a Challenge.

[e] Model based on generalized linear regression using negative binomial family and identity link, which was adjusted for Challenge and Tournament counts.

[f] N/A: not applicable.

[g] Model based on generalized linear regression using Poisson family and identity link, which was adjusted for Challenge and Tournament counts.

Significantly fewer website pages were viewed by the App Only (MD −69.5; 95% CI −72.9 to −66.1) and Web and App (MD −8.6; 95% CI −14.2 to −3.1) groups relative to the Website Only group, whereas significantly more pages were viewed by the Fitbit Only (MD 47.6; 95% CI 38.8 to 56.5) and Fitbit Combination (MD 50.8; 95% CI 42.5 to 59.0) groups. All groups, except the Fitbit Combination group, viewed significantly fewer website pages per session than the Website Only group.

The mean daily step count was significantly lower for the App Only (MD −201.9; 95% CI −387.7 to −116.0) and Fitbit Only (MD −492.9; 95% CI −679.9 to −305.8) groups relative to the Website Only group, whereas the mean daily step count of the Web and App group (MD 258.0; 95% CI 76.9- 439.2) was significantly higher. The App Only and Website Only groups did not significantly differ from each other in the number of step log entries, whereas the Fitbit Only, Web and App, and Fitbit Combination groups all logged steps a significantly greater number of times (MD 5.0, 95% CI 3.4-6.6; MD 7.2, 95% CI 5.9-8.6; and MD 15.6, 95% CI 13.7-17.5, respectively) than the Website Only group, with the Fitbit Combination group logging steps the highest number of times.

The App Only group was less likely (Challenges: OR 0.65, 95% CI 0.46-0.94; friend requests: OR 0.86, 95% CI 0.75-0.99) and all other groups were significantly more likely (Challenges: OR 2.38-4.48; *P*<.001; friend requests: OR 1.37-1.67; *P*<.001) than the Website Only group to participate in Challenges as well as send and/or receive friend requests. All groups were more likely than the Website Only group to participate in Tournaments (OR 1.26-1.79; *P*<.001; Table 3).

Table 3. Associations between participating in Challenges and Tournaments and receiving and sending friend requests from January 1 to December 31, 2018 among the 10,000 Step user groups.

User groups[a]	OR[b] (95% CI)	*P* value
Participating in individual Challenges[c] (reference category: no)		
Website Only	Reference group	N/A[d]
App Only	0.65 (0.46-0.94)	.02
Fitbit Only	3.29 (2.66-4.06)	<.001
Web and App	2.38 (1.91-2.97)	<.001
Fitbit Combination	4.48 (3.68-5.46)	<.001
Participating in team Tournaments[e] (reference category: no)		
Website Only	Reference group	N/A
App Only	1.26 (1.13-1.41)	<.001
Fitbit Only	1.28 (1.14-1.45)	<.001
Web and App	1.70 (1.51-1.91)	<.001
Fitbit Combination	1.79 (1.57-2.05)	<.001
Receiving and/or sending friend requests[f] (reference category: no)		
Website Only	Reference group	N/A
App Only	0.86 (0.75-0.99)	.04
Fitbit Only	1.67 (1.47-1.89)	<.001
Web and App	1.37 (1.21-1.54)	<.001
Fitbit Combination	1.44 (1.27-1.65)	<.001

[a] Website Only group: n=14,617; App Only group: n=2100; Fitbit Only group: n=1705; Web and App group: n=2057; Fitbit Combination group: n=1663.

[b] OR: odds ratio.

[c] Adjusted for Tournaments.

[d] N/A: not applicable.

[e] Adjusted for Challenges.

[f] Adjusted for Tournaments and Challenges.

Nonusage Attrition

Of the 22,142 new users who logged steps at least once, 21 (0.09%) did not succumb to nonusage attrition between the time of registration and December 31, 2018. The mean time to nonusage attrition was 35 (SD 26) days (Website Only: mean 32 days, SD 22 days; App Only: mean 33 days, SD 23 days; Fitbit Only: mean 40 days, SD 29 days; Web and App: mean 39 days, SD 27 days; Fitbit Combination: mean 50 days, SD 40 days). Among those who logged steps at least once, the estimated median time to nonusage (ie, the time after which 50% of users cease logging steps) was 31 days, with the Fitbit Combination group taking the longest time to reach this point (41 days; Table 4). All groups, except the App Only group, had a significant difference in time to nonusage attrition relative to the Website Only group (HR: range 0.55-0.75), and this association remained, although slightly attenuated, after adjusting for participation in Challenges and Tournaments (HR: range 0.59-0.78; Table 5; Figure 2, Figure 3). The Fitbit Combination group had a >40% lower likelihood of succumbing to nonusage attrition than the Website Only group.

Table 4. Cox proportional hazard risks for nonusage attrition of 10,000 Steps program, by user group and by participation in Challenges and Tournaments (N=30,040).

Group/participation	Unadjusted[a] HR[b] (95% CI)	*P* value	Adjusted[c] HR (95% CI)	*P* value
User group				
Website Only	Reference group	N/A[d]	N/A	N/A
App Only	0.97 (0.93-1.01)	.17	0.97 (0.93-1.02)	.22
Fitbit Only	0.71 (0.68-0.75)	<.001	0.75 (0.72-0.79)	<.001
Web and App	0.75 (0.72-0.79)	<.001	0.78 (0.74-0.81)	<.001
Fitbit Combination	0.55 (0.52-0.57)	<.001	0.59 (0.56-0.62)	<.001
Level of participation in Challenges and Tournaments				
Did not participate in Challenge	Reference group	N/A	N/A	N/A
Participated in Challenge	0.48 (0.45-0.52)	<.001	N/A	N/A
Did not participate in Tournament	Reference group	N/A	N/A	N/A
Participated in Tournament	0.75 (0.73-0.78)	<.001	N/A	N/A

[a] Unadjusted model.

[b] HR: hazard ratio.

[c] Model adjusted for participation in Challenges and Tournaments.

[d] N/A: not applicable.

Table 5. Survival time by group of users of the 10,000 Steps program (N=30,040).

Percentage of user groups still using the 10,000 Steps platform[a]	Duration (days)
75%	
Website Only	20
App Only	24
Fitbit Only	25
Web and App	29
Fitbit Combination	30
Nonloggers	0
All logging groups combined	22
All groups combined	0
50%	
Website Only	30
App Only	31
Fitbit Only	36
Web and App	33
Fitbit Combination	41
Nonloggers	0
All logging groups combined	31
All groups combined	27
25%	
Website Only	41
App Only	39
Fitbit Only	48
Web and App	44
Fitbit Combination	56
Nonloggers	0
All logging groups combined	43
All groups combined	37

[a] Participants may not have engaged with the program on all days consecutively.

Figure 2. Kaplan-Meier estimates of the survival distribution for time to nonusage attrition by group.

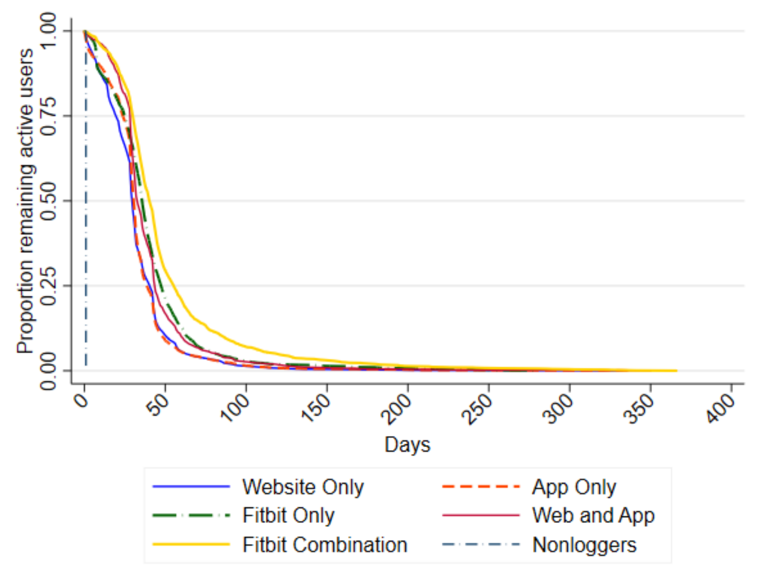
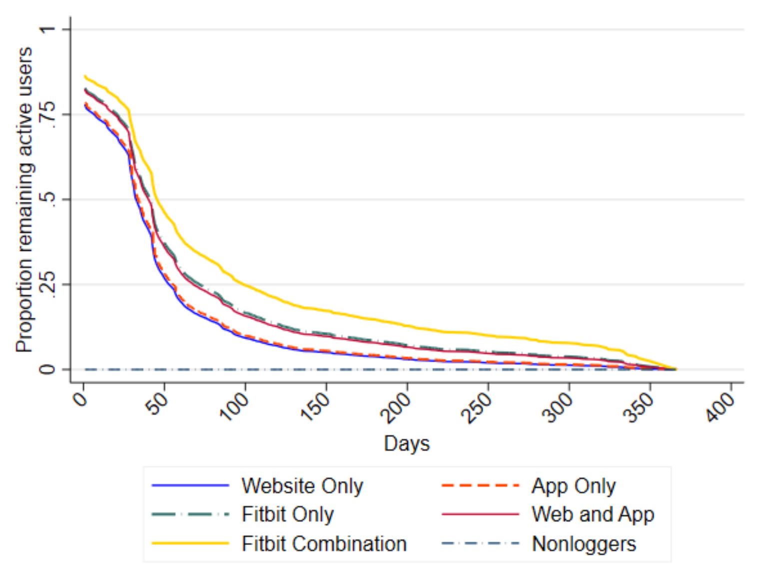


Figure 3. Cox proportional hazards regression curve of the survival distribution for time to nonusage attrition by group based on the model adjusted for Challenges and Tournaments; reference category: did not participate in Challenges or Tournaments.



Discussion

Principal Findings

This study examined whether different methods of logging steps in the 10,000 Steps physical activity program are associated with engagement with the program. The results showed that those who used Fitbits (Fitbit Only and Fitbit Combination) were the most engaged with the program. There was also some evidence that those who used the combination of Web and App were also well engaged with the program, but not to the same extent as Fitbit users.

The Fitbit Combination group had an engagement profile that included a high number of website sessions, page views, and step log entries as well as a higher likelihood of participating in both individual Challenges and team Tournaments. This group also had the lowest risk of succumbing to nonusage attrition, which corresponded with the longest time to succumb to nonusage attrition. The Fitbit Only group also had relatively good engagement characteristics and logged steps for the second longest time. These results suggest that the use of a Fitbit may enhance engagement with a real-world physical activity program. This also indicates that using a Fitbit in combination with other step-logging methods may further enhance engagement.

These findings were not expected because Fitbit users had the option to view their step counts either on the Fitbit itself or on the Fitbit website or app. Subsequently, they could have neglected to log in to the 10,000 Steps website to synchronize their Fitbit steps and interact with the other program content on the website. A number of factors may partially explain the superior engagement of the groups using Fitbits. The financial investment involved in purchasing a Fitbit may indicate that these device-owning individuals were more motivated to increase their activity levels compared with the other groups. Previous research has suggested that monetary investment may be associated with behavior change either because of the presence of motivation at the time of purchase or through financial investment stimulating motivation [26]. Advanced activity tracking devices have previously been shown to reduce the burden of self-monitoring activity compared with traditional self-monitoring methods such as pedometers [27]. The increased convenience may lead to improved engagement with other program features (eg, reading website content). In addition, wearing a Fitbit may serve as an activity-related prompt or cue. Another factor may include the nature of the Fitbits themselves. Users of activity tracking devices have cited a number of features that apply to Fitbits as being important characteristics, such as being wrist-worn; being accurate [28]; having the ability to synchronize with other devices; and having the ability to track additional items such as heart rate, distance traveled, and sleep [29,30]. Given that the level of engagement with activity trackers is associated with high user satisfaction [29] these factors may act as motivators for users to maintain program engagement and be active. Other features of Fitbits that may improve engagement within these groups include the associated Fitbit app, which may have increased exposure to additional motivational messaging and the incorporation of behavior change techniques such as goal setting and feedback [8]. Engagement may be active (ie, logging data or completing quizzes) or passive (viewing the intervention without interacting) [31], and it is unclear how Fitbit use may influence this as the data are automatically synchronized to the 10,000 Steps platform. This aspect of engagement with a Fitbit may be passive relative to the manual entry of data. Furthermore, differences in broader indicators of use, such as session duration and page views, were not consistent between the groups that included Fitbit. Consequently, it may be interesting for future studies to examine the mode of self-monitoring in the context of a broader set of engagement indicators.

The Fitbit Combination group was less likely to succumb to nonusage attrition and took longer to succumb to nonusage attrition. This might reflect research showing that behavior change and the sustained use of activity trackers are enhanced when trackers (such as Fitbit) that provide feedback are used in conjunction with other interventions (such as the 10,000 Steps program), which delineate target behaviors and provide a plan of action [32]. In addition, research exploring the relationship between the use of a Fitbit and changes in physical activity suggested that being accountable to someone else had a greater influence on increasing physical activity than simply self-monitoring the data on a Fitbit [27]. A sense of accountability might have been created through registration with the 10,000 Steps program and by the greater participation in individual Challenges and team Tournaments among the Fitbit Combination group. Furthermore, team Tournaments involving social interaction are associated with improved adherence to physical activity interventions [33].

A higher proportion of women were found in the 2 groups that used Fitbits. This is contrary to previous research showing that men were more likely to use advanced activity trackers [29]. Previous research also indicated that women were more likely to use any type of activity tracker [29]; therefore, perhaps the fast-growing adoption of advanced activity trackers is now greater in women. It has been shown that the sustained use of activity trackers is longer among women than in men [32], and this factor might have contributed to the longer time to nonusage among the Fitbit Combination group, which had the highest proportion of women.

The App Only group was significantly younger, with a higher proportion of men than the other groups. This group demonstrated the lowest overall interaction with program content and was among the groups with the lowest mean daily step count and the earliest time to nonusage attrition. These findings suggest that additional strategies may be required to better engage younger male users of the program. Interestingly, 2 previous studies of the 10,000 Steps programs, which also examined app use, found contrasting outcomes [18,34]. The first study found the App Only group to be the youngest but mostly female, with engagement and time to nonusage attrition being better than those in the Website Only group [18]. The second study, which was undertaken when the app was first introduced, examined the app and web users of the 10,000 Steps program and found that app users logged more steps more often than Website Only users [34]. These discrepancies in findings could potentially be explained by those users who tend to be the *early adopters* of innovations [35]: those who were *early adopters* of the app when it was first introduced may have similar characteristics to those who are now the *early adopters* of the Fitbit functionality.

The groups that used a combination of step-logging methods had the best program engagement profile. This is highlighted by the greater number of step log entries. The higher the mean daily step count, the greater the likelihood of participating in individual Challenges and team Tournaments, and the longer the time until nonusage attrition by the Fitbit Combination group compared with the Fitbit Only group and by the Web and App group compared with both the App Only and Website Only groups. This consolidates the findings from a previous study of engagement with the 10,000 Steps program that found the Web and App group to have a better engagement profile than the Website Only and App Only groups [18]. It is difficult to determine whether the act of using multiple methods to log steps maintains a longer interest in the program or whether it is a personality type or level of motivation to be active, which leads them to both use several logging methods and stay engaged.

Nonusage attrition is a common problem among internet-delivered health interventions [36]. The mean time to nonusage attrition in this sample was similar to previous studies of the 10,000 Steps program, and both studies found that groups that logged steps using multiple methods (eg, using a combination of Web and App to log steps) took longer to cease logging steps [18]. The time to nonusage attrition in this study also compared favorably with several other mobile health physical activity studies conducted under both real-world settings (35 days in this study vs 11 days in another study [37]) and tightly controlled conditions (35 days in this study vs 32 [38] and 46 days [39] in other studies).

The mean daily step counts of all the groups was greater than 10,000 Steps, which is above the average of 7400 for Australian adults [40]. Although this is a promising indicator that this real-world physical activity intervention is effective, the magnitude of differences between the groups was <500 per day. Consequently, the broader public health implications of these differences are unclear and may be more relevant and meaningful for less active people, given that they benefit most from small increases in activity [41]. Alternatively, it may reflect that more active individuals are attracted to the program. The mean daily step counts were also the highest for the groups that used more than one method to log steps. However, the Fitbit Only group took the fewest steps, including fewer steps than the Website Only and App Only groups, despite the groups' other engagement parameters being generally better. This might also be indicative of the different impacts on behavior stemming from active versus passive self-monitoring [27]. Those required to manually enter their daily step counts may be more likely to consciously scrutinize their step counts during the process and consequently make a greater effort to take more steps. Meanwhile, the passive nature of the automatic synchronization of step counts by the Fitbit, when not linked to other interactions with program content, may lead to a loss of attention to step counts and subsequently to less stimulus to change behavior. A previous study of Fitbit users found that there were better improvements in physical activity among users who interacted with both the app and the Fitbit device than those who just checked their device [27]. Therefore, it is possible that the groups in this study who used a greater number of methods to monitor their steps might be undertaking active rather than passive self-monitoring (ie, paying attention to the feedback or graphs, etc), which led to better engagement by these groups.

Strengths and Limitations

Among the strengths of this study is the real-world delivery of a web-based physical activity intervention, not in a controlled setting. In addition, the sample was large and examined over a long time frame. The findings of this type of study are likely to provide more accurate information regarding how interventions work when they are delivered in ecologically valid settings [6]. Other strengths include the examination of various combinations of user step-logging methods that allow for a nuanced understanding of user engagement patterns. There are several limitations to consider. The results of this study must be interpreted in the context of the vast majority of users belonging to the Website Only group, and the outcomes may have been different if the groups were more evenly balanced. In addition, the only measure of physical activity was step count, which does not necessarily capture the overall physical activity. Furthermore, the Fitbit groups provided objective step data that were automatically synced with the website, whereas the others provided manually entered pedometer measured step counts that are prone to bias (ie, people may report more than what the pedometer actually measured). This might have created a disparity in the accuracy of step counts between the groups. Only 25% (5536/22,142) of those who actively logged steps were still engaged after 43 days; therefore, it is unknown whether the impressive >10,000 daily mean step counts of all groups was ongoing. It is noteworthy that the engagement metrics used in this study did not include broader indicators of different types of engagement (reflective, altruistic, and gamified), which may be important to behavior change [31].

Conclusions

This study found that multiple methods of logging steps were associated with better program engagement. The use of a Fitbit appears to enhance engagement with a real-world physical activity program, particularly when used in conjunction with other platforms (ie, a combination of Fitbit and website and/or app). Therefore, integrating tracking devices that synchronize data automatically into real-world physical activity interventions is one strategy to improve engagement.

Acknowledgments

MJD is supported by a Career Development Fellowship (APP1141606) from the National Health and Medical Research Council. CV was supported by a Future Leader Fellowship (ID 100427) from the National Heart Foundation of Australia. The 10,000 Steps program and the social media campaign were funded by the State Health Department (Queensland Health) of the Queensland Government.

Conflicts of Interest

None declared.

References

1. Lee I, Shiroma E, Lobelo F, Puska P, Blair S, Katzmarzyk P, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
2. Warburton D, Bredin S. Reflections on physical activity and health: what should we recommend? Can J Cardiol 2016 Apr;32(4):495-504. [doi: 10.1016/j.cjca.2016.01.024] [Medline: 26995692]
3. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, Lancet Physical Activity Series 2 Executive Committee. Progress in physical activity over the Olympic quadrennium. Lancet 2016 Sep 24;388(10051):1325-1336. [doi: 10.1016/S0140-6736(16)30581-5] [Medline: 27475270]
4. Bennie JA, Pedisic Z, van Uffelen JG, Gale J, Banting LK, Vergeer I, et al. The descriptive epidemiology of total physical activity, muscle-strengthening exercises and sedentary behaviour among Australian adults--results from the National

Nutrition and Physical Activity Survey. BMC Public Health 2016 Jan 25;16:73 [FREE Full text] [doi: 10.1186/s12889-016-2736-3] [Medline: 26809451]

5. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of ehealth and mhealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-28.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
6. Vandelanotte C, Duncan MJ, Kolt GS, Caperchione CM, Savage TN, Van Itallie A, et al. More real-world trials are needed to establish if web-based physical activity interventions are effective. Br J Sports Med 2019 Dec 3;53(24):1553-1554. [doi: 10.1136/bjsports-2018-099437] [Medline: 29970409]
7. Reis RS, Salvo D, Ogilvie D, Lambert EV, Goenka S, Brownson RC, Lancet Physical Activity Series 2 Executive Committee. Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. Lancet 2016 Sep 24;388(10051):1337-1348 [FREE Full text] [doi: 10.1016/S0140-6736(16)30728-0] [Medline: 27475273]
8. 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]
9. Bravata D, Smith-Spangler C, Sundaram V, Gienger A, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
10. Graph C. Unit Sales of Wearables Worldwide From 2014 to 2017 (in Millions). PublishedStatistica. 2018. URL: https://www.statista.com/statistics/757623/wearables-sales-by-category-worldwide/ [accessed 2019-02-01]
11. Percentage of the Global Population That Used a Mobile App or Fitness Tracking Device to Track Their Health as of 2016. Age Statistica. 2017. URL: https://www.statista.com/statistics/742448/global-fitness-tracking-and-technology-by-age/ [accessed 2019-02-01]
12. Graph C. Wearable Fitness Device Adoption Rate Worldwide in 2016. Statistica. 2019. URL: https://www.statista.com/statistics/651681/wearable-fitness-device-adoption-by-country/ [accessed 2019-02-01]
13. Alley S, Jennings C, Plotnikoff RC, Vandelanotte C. An evaluation of web- and print-based methods to attract people to a physical activity intervention. JMIR Res Protoc 2016 May 27;5(2):e94 [FREE Full text] [doi: 10.2196/resprot.4826] [Medline: 27235075]
14. Vandelanotte C, Duncan M, Maher C, Schoeppe S, Rebar A, Power D, et al. The effectiveness of a web-based computer-tailored physical activity intervention using FitBit activity trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [FREE Full text] [doi: 10.2196/11321] [Medline: 30563808]
15. Duncan MJ, Vandelanotte C, Trost SG, Rebar AL, Rogers N, Burton NW, et al. Balanced: a randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults. BMC Public Health 2016 Jul 30;16:670 [FREE Full text] [doi: 10.1186/s12889-016-3256-x] [Medline: 27473327]
16. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
17. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
18. Guertler D, Vandelanotte C, Kirwan M, Duncan MJ. Engagement and nonusage attrition with a free physical activity promotion program: the case of 10,000 steps Australia. J Med Internet Res 2015 Jul 15;17(7):e176 [FREE Full text] [doi: 10.2196/jmir.4339] [Medline: 26180040]
19. 10,000 Steps. URL: https://www.10000steps.org.au/ [accessed 2019-08-01]
20. Duncan MJ, Brown WJ, Mummery WK, Vandelanotte C. 10,000 steps Australia: a community-wide eHealth physical activity promotion programme. Br J Sports Med 2018 Jul 4;52(14):885-886. [doi: 10.1136/bjsports-2017-097625] [Medline: 28870978]
21. Brown WJ, Eakin E, Mummery K, Trost SG. 10,000 steps Rockhampton: establishing a multi-strategy physical activity promotion project in a community. Health Promot J Aust 2003 Aug 1;14(2):95-100. [doi: 10.1071/he03095]
22. Vandelanotte C, van Itallie A, Brown W, Mummery W, Duncan M. Every step counts: understanding the success of implementing the 10,000 steps project. Stud Health Technol Inform 2020 Mar 2;268:15-30. [doi: 10.3233/SHTI200003] [Medline: 32141876]
23. Kolt G, Rosenkranz R, Vandelanotte C, Caperchione C, Maeder A, Tague R, et al. Using web 2.0 applications to promote health-related physical activity: findings from the WALK 2.0 randomised controlled trial. Br J Sports Med 2017 Oct;51(19):1433-1440 [FREE Full text] [doi: 10.1136/bjsports-2016-096890] [Medline: 28049624]
24. Computer Program. Stata Statistical Software. 2017. URL: https://www.stata.com/support/faqs/resources/citing-software-documentation-faqs/ [accessed 2019-08-01]
25. Davies C, Corry K, van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]

26. Sullivan AN, Lachman ME. Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. Front Public Health 2016;4:289 [FREE Full text] [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]
27. Hartman SJ, Nelson SH, Weiner LS. Patterns of fitbit use and activity levels throughout a physical activity intervention: exploratory analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2018 Feb 5;6(2):e29 [FREE Full text] [doi: 10.2196/mhealth.8503] [Medline: 29402761]
28. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
29. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. BMJ Open 2016 Jul 7;6(7):e011243 [FREE Full text] [doi: 10.1136/bmjopen-2016-011243] [Medline: 27388359]
30. Coughlin S, Stewart J. Use of consumer wearable devices to promote physical activity: a review of health intervention studies. J Environ Health Sci 2016 Nov;2(6):- [FREE Full text] [doi: 10.15436/2378-6841.16.1123] [Medline: 28428979]
31. 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]
32. Hermsen S, Moons J, Kerkhof P, Wiekens C, De GM. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
33. DiMatteo MR. Social support and patient adherence to medical treatment: a meta-analysis. Health Psychol 2004 Mar;23(2):207-218. [doi: 10.1037/0278-6133.23.2.207] [Medline: 15008666]
34. Kirwan M, Duncan MJ, Vandelanotte C, Mummery WK. Using smartphone technology to monitor physical activity in the 10,000 Steps program: a matched case-control trial. J Med Internet Res 2012 Apr 20;14(2):e55 [FREE Full text] [doi: 10.2196/jmir.1950] [Medline: 22522112]
35. Dearing JW, Cox JG. Diffusion of innovations theory, principles, and practice. Health Aff (Millwood) 2018 Feb;37(2):183-190. [doi: 10.1377/hlthaff.2017.1104] [Medline: 29401011]
36. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
37. Vandelanotte C, Kolt G, Caperchione C, Savage T, Rosenkranz R, Maeder A, et al. Effectiveness of a web 2.0 intervention to increase physical activity in real-world settings: randomized ecological trial. J Med Internet Res 2017 Nov 13;19(11):e390 [FREE Full text] [doi: 10.2196/jmir.8484] [Medline: 29133282]
38. Murawski B, Plotnikoff R, Rayward A, Oldmeadow C, Vandelanotte C, Brown W, et al. Efficacy of an m-health physical activity and sleep health intervention for adults: a randomized waitlist-controlled trial. Am J Prev Med 2019 Oct;57(4):503-514. [doi: 10.1016/j.amepre.2019.05.009] [Medline: 31542128]
39. Rayward A, Murawski B, Duncan M, Holliday E, Vandelanotte C, Brown W, et al. Efficacy of an m-health physical activity and sleep intervention to improve sleep quality in middle-aged adults: the refresh study randomized controlled trial. Ann Behav Med 2020 Jun 12;54(7):470-483. [doi: 10.1093/abm/kaz064] [Medline: 31942918]
40. Australian Health Survey: Physical Activity. Australian Bureau of Statistics. 2020. URL: https://beta.abs.gov.au/statistics/health/health-conditions-and-risks/australian-health-survey-physical-activity/2011-2012 [accessed 2020-01-18]
41. Oftedal S, Holliday E, Attia J, Brown W, Collins C, Ewald B, et al. Daily steps and diet, but not sleep, are related to mortality in older Australians. J Sci Med Sport 2020 Mar;23(3):276-282. [doi: 10.1016/j.jsams.2019.09.018] [Medline: 31615727]

Abbreviations

HR: hazard ratio
MD: mean difference
OR: odds ratio

Edited by R Kukafka; submitted 06.07.20; peer-reviewed by T Raedeke, R Jiwani; comments to author 21.08.20; revised version received 16.10.20; accepted 25.02.21; published 18.06.21.

Please cite as:
Rayward AT, Vandelanotte C, Van Itallie A, Duncan MJ
The Association Between Logging Steps Using a Website, App, or Fitbit and Engaging With the 10,000 Steps Physical Activity Program: Observational Study
J Med Internet Res 2021;23(6):e22151
URL: https://www.jmir.org/2021/6/e22151
doi:10.2196/22151
PMID:34142966

Original Paper

Association of Habitual Physical Activity With Home Blood Pressure in the Electronic Framingham Heart Study (eFHS): Cross-sectional Study

Mayank Sardana[1*], MBBS, MSc; Honghuang Lin[2*], PhD; Yuankai Zhang[3], BSc; Chunyu Liu[3], PhD; Ludovic Trinquart[3], PhD, MPH; Emelia J Benjamin[2,3,4], MD, ScM; Emily S Manders[4], BSc; Kelsey Fusco[4], BSc; Jelena Kornej[4], MSc, MD; Michael M Hammond[4], MD; Nicole Spartano[2], PhD; Chathurangi H Pathiravasan[3], PhD; Vik Kheterpal[5], MD; Christopher Nowak[5], PhD; Belinda Borrelli[6], PhD; Joanne M Murabito[2,4], MD, ScM; David D McManus[7], MD, ScM

[1]Department of Medicine, Division of Cardiology, University of California San Francisco, San Francisco, CA, United States
[2]Department of Medicine, Boston University School of Medicine, Boston, MA, United States
[3]Boston University School of Public Health, Boston, MA, United States
[4]Framingham Heart Study, Framingham, MA, United States
[5]Care Evolution, Ann Arbor, MI, United States
[6]Henry M Goldman School of Dental Medicine, Center for Behavioral Science Research, Boston University, Boston, MA, United States
[7]Department of Medicine, UMass Medical School, Worcester, MA, United States
[*]these authors contributed equally

Corresponding Author:
Mayank Sardana, MBBS, MSc
Department of Medicine, Division of Cardiology
University of California San Francisco
505 Parnassus Ave
San Francisco, CA, 94143
United States
Phone: 1 4157548491
Email: mayank.mamc@gmail.com

Abstract

Background: When studied in community-based samples, the association of physical activity with blood pressure (BP) remains controversial and is perhaps dependent on the intensity of physical activity. Prior studies have not explored the association of smartwatch-measured physical activity with home BP.

Objective: We aimed to study the association of habitual physical activity with home BP.

Methods: Consenting electronic Framingham Heart Study (eFHS) participants were provided with a study smartwatch (Apple Watch Series 0) and Bluetooth-enabled home BP cuff. Participants were instructed to wear the watch daily and transmit BP values weekly. We measured habitual physical activity as the average daily step count determined by the smartwatch. We estimated the cross-sectional association between physical activity and average home BP using linear mixed effects models adjusting for age, sex, wear time, antihypertensive drug use, and familial structure.

Results: We studied 660 eFHS participants (mean age 53 years, SD 9 years; 387 [58.6%] women; 602 [91.2%] White) who wore the smartwatch 5 or more hours per day for 30 or more days and transmitted three or more BP readings. The mean daily step count was 7595 (SD 2718). The mean home systolic and diastolic BP (mmHg) were 122 (SD 12) and 76 (SD 8). Every 1000 increase in the step count was associated with a 0.49 mmHg lower home systolic BP (*P*=.004) and 0.36 mmHg lower home diastolic BP (*P*=.003). The association, however, was attenuated and became statistically nonsignificant with further adjustment for BMI.

Conclusions: In this community-based sample of adults, higher daily habitual physical activity measured by a smartwatch was associated with a moderate, but statistically significant, reduction in home BP. Differences in BMI among study participants accounted for the majority of the observed association.

(J Med Internet Res 2021;23(6):e25591) doi:10.2196/25591

KEYWORDS

hypertension; primary prevention; eCohort; physical activity; smartwatch; Apple Watch; home blood pressure

Introduction

Understanding the relationship between physical activity and blood pressure (BP) is crucial because promoting physical activity might help address the community burden of hypertension. Several observational and interventional studies have explored the association between physical activity and BP. Data from community-based observational studies suggest that higher self-reported moderate to vigorous physical activity, but not overall physical activity, is associated with lower research center/in-office BP and lower prevalence of incident hypertension [1-6]. Similarly, in interventional studies, promoting moderate to vigorous physical activity leads to a consistent reduction in BP in both normotensive and hypertensive individuals [7,8]. Most individuals, however, spend the majority of their time performing light activities, such as walking, which are suboptimally quantified using physical activity questionnaires [9,10], but can be accurately measured using accelerometers or smartwatches. Understanding the relationship between habitual physical activity and BP might yield fruitful targets to address the community burden of hypertension.

Smartwatches are commercially available devices for monitoring habitual physical activity and can enhance phenotyping of community-dwelling individuals. Daily step count reported by smartwatches could be a useful measure of overall physical activity capturing all intensities of exercise. Home BP, when compared to in-office BP, is a stronger predictor of adverse cardiovascular outcomes and mortality [11]. Merging the vast data continuously being collected using smart gadgets, such as smartwatches and Bluetooth-enabled BP cuffs, not only enriches the phenotypical information for individuals, but also opens up the possibility of studying the relationship of these novel ambulatory phenotypes (eg, daily step count) with cardiovascular risk factors (such as home BP).

We hypothesized that higher habitual physical activity is associated with lower home BP. To that end, we leveraged data from the ongoing electronic Framingham Heart Study (eFHS) cohort [12] to assess the association between home BP and habitual physical activity measured using a smartwatch (daily step count). Additionally, we sought to study the moderators of this association by adjusting for previously known correlates of hypertension and physical activity.

Methods

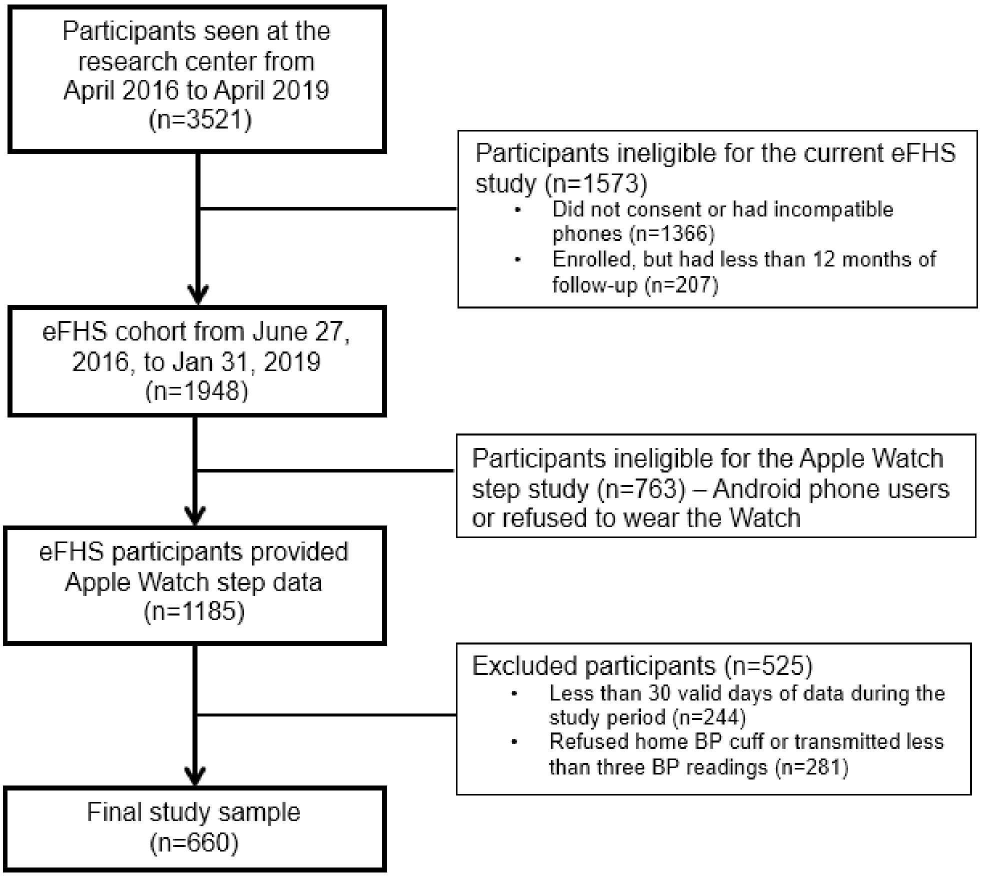
Overview

The Framingham Heart Study (FHS) is a multigenerational cohort study that was originally designed to study the risk factors for cardiovascular disease [13]. The eFHS cohort started enrolling participants from the FHS Third Generation Cohort (Gen 3), multiethnic Omni Group 2 Cohort (Omni 2), and New Offspring Spouse Cohort in June 2016 during regular research center examination [12]. The participants were offered a smartphone app, a Nokia Withings BP cuff for home BP monitoring, and a smartwatch (Apple Watch, Series 0, started November 2016). To be eligible for the BP cuff and Apple Watch, the participants were required to own an iPhone with a compatible iOS version (version 9 or higher). We chose the Withings wireless BP device because it is Food and Drug Administration approved for home BP monitoring and it has been validated (mean differences between the device and mercury readings for systolic and diastolic BP of −0.2 [SD 5.0] mmHg and 0.4 [SD 4.2] mmHg, respectively) [14,15]. The research protocol was approved by the Boston University Medical Center Institutional Review Board. All participants provided written informed consent.

Study Protocol

In this study, eFHS participants enrolled between June 27, 2016, and January 31, 2019, were included, and the median follow-up was 375 days from the date of enrollment (Figure 1). The participants were asked to wear the Apple Watch every day and transmit home BP readings every week. Habitual physical activity was measured as the average daily step count transmitted by the smartwatch. To reduce the bias related to low watch wear time, we only averaged the data from "valid days," which were defined as those days with at least 5 hours of wear time. We, however, also performed sensitivity analyses leveraging an alternate threshold of 10 hours/day to define valid days. Of note, wear time from the Apple Watch was determined based on a combination of the heart rate and step count data transmitted by the watch. At rest, the watch measures the heart rate opportunistically every 5 minutes (heart rate is measured at a higher frequency during workouts) [16]. Therefore, any clock hours with more than one heart rate recording were included in the wear time. For the hours with one or less heart rate recordings, we only included clock hours with 30 or more steps in the wear time. Additionally, participants with less than 30 valid days were excluded (n=244; Figure 1). Several prior studies have measured physical activity by deploying accelerometers over short periods of time (eg, 1 week), which carries the risk of potential bias, as the participants might modify their behavior for the duration of the study [7,17]. In comparison to accelerometer-based studies, we leveraged physical activity measured by the smartwatch over a median follow-up of just over a year and did not include the data from participants with less than 30 valid days. Therefore, we likely captured the "habitual" level of physical activity, that is, the level of physical activity a participant will automatically perform if not being monitored (eg, using accelerometers).

Figure 1. Cohort development diagram for the study. BP: blood pressure; eFHS: electronic Framingham Heart Study.



Home BP Measurement

The participants were advised to measure home BP once a week, ideally on the same day every week and at the same time during the day. They were advised to sit in a comfortable position with the feet flat on the floor and the left arm resting on a table with the palm up, rest for 5 minutes without talking, and then take the BP measurement. Participants were advised against BP measurement after exercise, after consuming caffeinated beverages, or after a high excitement activity. All BP recordings were date and time stamped. The BP recordings taken by the participants during the study period were averaged to calculate average home systolic and diastolic BP for each participant. We only included participants who transmitted three or more home BP recordings. We chose a threshold of three readings because prior studies have reported that home BP readings averaged over 3 days versus 7 or 10 days yield similar results [18,19]. We, however, did perform sensitivity analyses leveraging an alternate threshold of nine or more home BP readings (as described below). The variability in home BP readings was measured using the coefficient of variation (standard deviation/mean).

Other Variables

Clinical and laboratory variables were measured during the examination at the research center [13]. Hypertension was defined as systolic BP ≥140 mmHg, diastolic BP ≥90 mmHg, and/or self-reported use of antihypertensive medications. Diabetes was defined as fasting plasma glucose ≥126 mg/dL and/or self-reported use of medications for diabetes. BMI was calculated by dividing body weight (in kg) by height (in meters) square.

Statistical Methods

Baseline variables (from research center examination) were presented as mean (SD) or median (IQR) for continuous variables and as frequency (proportion) for categorical variables. The correlation between average daily step count and home systolic and diastolic BP was determined using the Spearman correlation coefficient and depicted using a correlation matrix. The association between average daily step count and home BP was measured using separate linear mixed effect regression models for systolic and diastolic BP. Primary models were adjusted for age, sex, familial structure/relatedness in the FHS, antihypertensive drug use, and watch wear time. The secondary model was further adjusted for BMI. We checked for interaction with wear time using the interaction term step count×wear time. For each model, we also performed sex-stratified analyses and tested assumption of linearity by inspecting the residual plots. As the FHS is a multigenerational study, to account for relatedness between the participants, we included the familial relatedness variable. This variable was derived from the self-reported pedigree structure by R package kinship and was treated as a random effect in our regression models. Additionally, sensitivity analyses were performed leveraging alternate inclusion thresholds for valid days (60 days or 90 days instead of 30 days), wear time (10 hours per valid day vs 5 hours per valid day), and home BP recordings (minimum nine home BP recordings vs three recordings). Similarly, we studied the association between home BP variability and average daily step count in primary and secondary models. Additionally, BMI-stratified exploratory analyses were performed to study the association between home BP and daily step count across various BMI strata (normal BMI <25 kg/m^2, overweight BMI 25-29.9 kg/m^2, and obese BMI ≥30 kg/m^2). These stratified models were adjusted for age, sex, familial structure, antihypertensive drug use, and watch wear time (similar to the primary models). We also performed exploratory analyses using regression models 1 and 2 in a subsample of participants with a history of hypertension. To account for multiple comparisons in these exploratory post-hoc stratified analyses, Bonferroni correction for the *P* value was used to determine statistical significance (*P*<.05/3=.017). For other a priori models, the significant association was defined by a two-sided *P* value <.05. Box plots were used to show the distribution of average daily step count and home BP across BMI strata. All statistical analyses were performed using R software package version 3.5.0 (The R Project for Statistical Computing).

Results

Our study sample consisted of 660 adults (mean age 53 years, SD 9 years; 387 [58.6%] women; 602 [91.2%] White; Table 1). Approximately one in five participants reported taking antihypertensive medications. The mean BMI was 27.8 kg/m^2 (SD 5 kg/m^2), and over two-thirds of participants were overweight (n=277) or obese (n=183). The baseline prevalence of smoking, diabetes, and cardiovascular disease was low. In Multimedia Appendix 1, we have presented the baseline characteristics of the study participants including all research center exam attendees and eFHS cohort participants. The eFHS participants were younger and more likely to be female, compared to all research center examinees. Other baseline characteristics were comparable between all eFHS participants and the study participants.

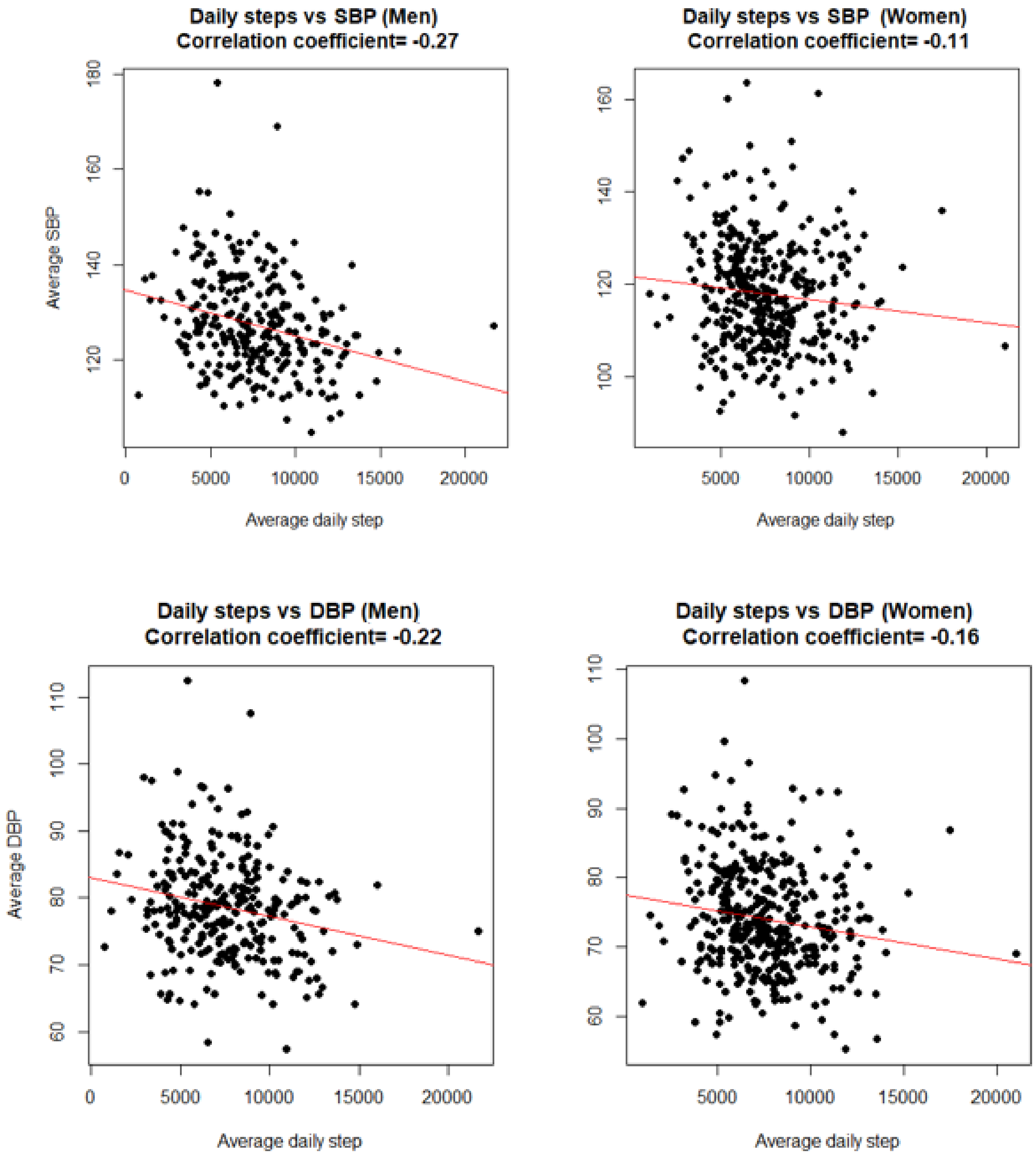
Table 1. Baseline characteristics of the study participants (N=660).

Variable	Value, mean (SD) or n (%)
Age (years), mean (SD)	53 (9)
Female sex, n (%)	387 (58.6%)
BMI (kg/m^2), mean (SD)	27.8 (5.0)
Systolic BP[a] (research center) (mmHg), mean (SD)	119 (14)
Diastolic BP (research center) (mmHg), mean (SD)	76 (9)
Antihypertensive drug use, n (%)	145 (22.0%)
Current smoking, n (%)	28 (4.2%)
Diabetes mellitus, n (%)	41 (6.2%)
Race, n (%)	
White	602 (91.2%)
Black	14 (2.1%)
Hispanic	17 (2.6%)
Asian	13 (2.0%)
Other	14 (2.1%)
Cardiovascular disease, n (%)	26 (3.9%)
Systolic BP (home) (mmHg), mean (SD)	122 (12)
Diastolic BP (home) (mmHg), mean (SD)	76 (8)
Daily step count, mean (SD)	7595 (2718)

[a] BP: blood pressure.

Over a median of 375 (25%-75%: 180-581) follow-up days, participants wore the Apple Watch for a median of 13.7 (IQR 12.4-14.8) hours per day and sent BP readings for 28 (IQR 11-63) weeks. The mean daily step count was 7595 (SD 2718). The mean home systolic and diastolic BP (mmHg) values were 122 (SD 12) and 76 (SD 8), respectively. In sex-stratified analyses, average daily step count was inversely correlated with home systolic and diastolic BP (in both men and women; Figure 2), that is, participants with a higher daily step count had lower home systolic and diastolic BP.

Figure 2. Scatter plots depicting the correlation of average daily step count with average home systolic and diastolic blood pressure. DBP: diastolic blood pressure; SBP: systolic blood pressure.



In the regression models adjusted for age, sex, watch wear time, family structure, and antihypertensive drug use (model 1), each 1000-step increment was associated with a 0.49 mmHg lower home systolic BP and 0.36 mmHg lower home diastolic BP (Table 2, Figure 3). No significant interaction with wear time was noted (*P*=.82 for systolic BP analyses and *P*=.35 for diastolic BP analyses). In sex-stratified analyses, the findings were overall similar in men and women, although the *P* value did not reach the statistical significance threshold in men for diastolic BP. Given the same directionality of the association in men as the overall sample and subsample with women and a relatively weak correlation between diastolic BP and daily steps (correlation coefficient −0.22), we suspected that we were underpowered in our sex-stratified analyses. We have included the power calculation in Multimedia Appendix 2. With further adjustment for BMI (model 2), the strength of the association between daily step count and home BP attenuated and became statistically nonsignificant. When we studied the association of log-transformed daily step count with home BP, the results remained unchanged (Multimedia Appendix 3).

Table 2. Association of daily step count with home blood pressure using separate mixed linear effect models for systolic and diastolic blood pressure.

Home BP[a] and participants	Model 1[b]			Model 2[c]		
	β[d] (mmHg)	SE	*P* value	β[d] (mmHg)	SE	*P* value
Systolic BP						
All participants (n=660)	−0.49	0.17	.004	0.006	0.16	.97
Women (n=387)	−0.47	0.23	.05	0.08	0.20	.70
Men (n=273)	−0.53	0.24	.03	−0.14	0.24	.55
Diastolic BP						
All participants (n=660)	−0.36	0.12	.003	−0.04	0.12	.76
Women (n=387)	−0.45	0.16	.01	−0.08	0.14	.59
Men (n=273)	−0.26	0.19	.18	−0.01	0.19	.94

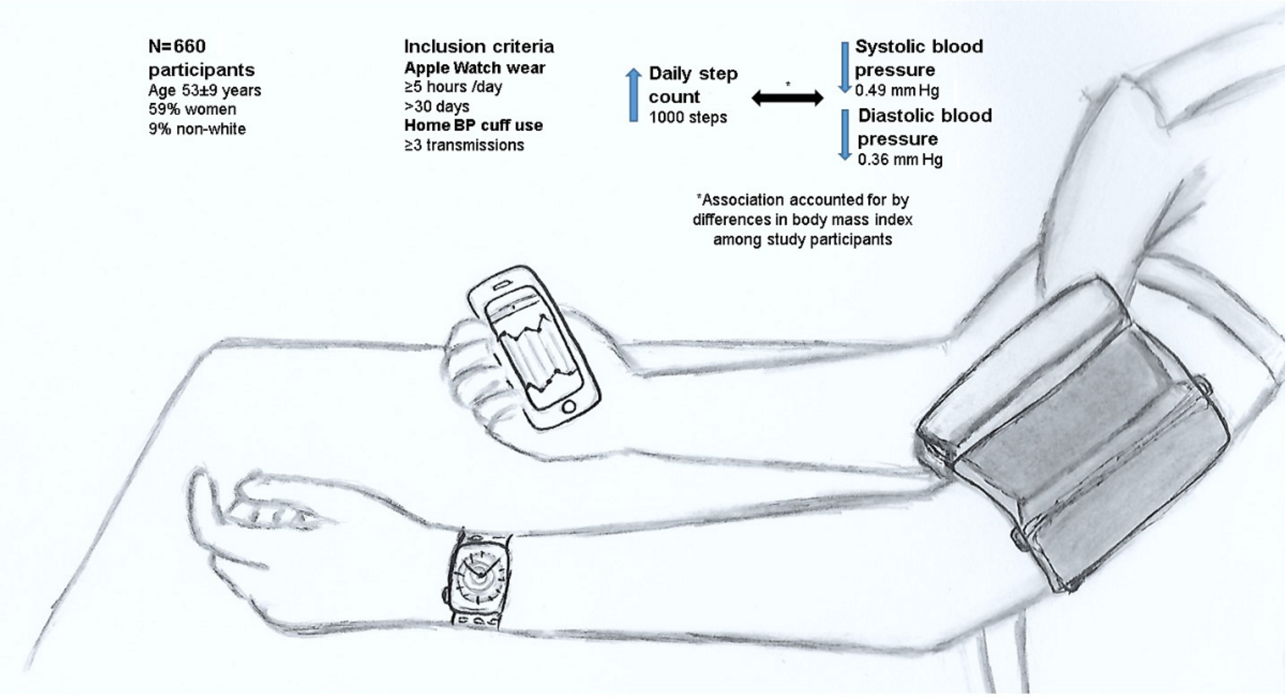
[a] BP: blood pressure.

[b] Model 1 was adjusted for age, sex, family structure, reported antihypertensive drug use, and watch wear time.

[c] Model 2 was adjusted for model 1 covariates and BMI.

[d] β represents the change in BP (mmHg) for every 1000 increase in daily steps.

Figure 3. Association of daily step count with home blood pressure. BP: blood pressure.



Of note, our primary analyses were restricted to the participants who had at least 30 active days and transmitted three or more home BP readings. We performed four sensitivity analyses to test the generalizability of our findings and increase the validity of our results as follows: (1) participants with 60 or more active days (n=611; Multimedia Appendix 4), (2) participants with 90 or more active days (n=578; Multimedia Appendix 5), (3) participants with nine or more home BP readings (n=540; Multimedia Appendix 6), and (4) using a 10 hours per valid day threshold (n=634; Multimedia Appendix 7). The results in sensitivity analyses were similar to the primary analyses, that is, there was an inverse association between daily step count and home BP in model 1, but the association became nonsignificant in model 2.

We also studied the association of home BP variability (coefficient of variation) with the daily step count using separate mixed linear effect models for systolic BP and diastolic BP in models analogous to the primary analyses (model 1 adjusted for age, sex, watch wear time, antihypertensive drug use, and family structure, and model 2 further adjusted for BMI). In these models, a higher daily step count was associated with significantly lower diastolic BP variability (but not systolic BP variability; Table 3). In sex-stratified analyses, the findings were overall similar in men and women, although the *P* value did not reach the statistical significance threshold in women for diastolic BP variability. Given the same directionality of the association in women as the overall sample and subsample with men, we suspected that we were underpowered in our sex-stratified analyses.

To further explore the association of BMI with home BP and daily step count, we performed BMI-stratified analyses. In these models adjusted for age, sex, familial structure, antihypertensive drug use, and watch wear time, we did not observe any statistically significant association of home BP (separate models for systolic and diastolic BP) with daily step count (Table 4). In Multimedia Appendix 8 and Multimedia Appendix 9, we depict the distribution of average daily step count and home BP across different BMI strata (normal weight, overweight, and obese). There was a significant trend toward a lower step count and higher systolic and diastolic BP with a higher BMI stratum (*P*<.001). In Multimedia Appendix 10, we have shown scatter plots depicting the correlation of average daily step count and home BP stratified by different BMI categories. We also performed exploratory analyses among participants with a history of hypertension (n=183). Compared to participants without a history of hypertension, participants with a history of hypertension had a lower daily step count, were older, were more likely to be men, and had a higher BMI (Multimedia Appendix 11). In these analyses, we did not observe a significant association between step count and systolic or diastolic BP in models 1 and 2 (Multimedia Appendix 12). Although the findings might be because of underpowering, they suggest that the relationship of step count with home BP is not stronger among participants with a history of hypertension. When we treated home BP as a tertile variable, we observed results similar to those in our primary analyses (Multimedia Appendix 13).

Table 3. Association of daily step count with home blood pressure variability (coefficient of variation).

Home BP[a] and participants	Model 1[b]			Model 2[c]		
	β[d] (mmHg)	SE	*P* value	β[d] (mmHg)	SE	*P* value
Systolic BP						
All participants (n=660)	−0.0002	0.0003	.65	−0.00017	0.00034	.62
Women (n=387)	0.0001	0.0004	.72	0.00018	0.00042	.67
Men (n=273)	−0.0006	0.0005	.25	−0.00074	0.00057	.19
Diastolic BP						
All participants (n=660)	−0.0011	0.0004	.003	−0.00095	0.00037	.01
Women (n=387)	−0.0008	0.0004	.09	−0.00060	0.00045	.19
Men (n=273)	−0.0016	0.0006	.01	−0.00161	0.00064	.01

[a] BP: blood pressure.

[b] Model 1 was adjusted for age, sex, family structure, reported antihypertensive drug use, and watch wear time.

[c] Model 2 was adjusted for model 1 covariates and BMI.

[d] β represents the change in BP (mmHg) for every 1000 increase in daily steps.

Table 4. Association of daily step count with home blood pressure in models stratified by BMI.

Home BP[a] and participants	Normal (BMI <25 kg/m^2)[b] (n=200)			Overweight (25 ≤ BMI < 30 kg/m^2)[b] (n=277)			Obese (BMI ≥30 kg/m^2)[b] (n=183)		
	β[c] (mmHg)	SE	*P* value	β[c] (mmHg)	SE	*P* value	β[c] (mmHg)	SE	*P* value
Systolic BP									
All participants	−0.04	0.25	.88	−0.34	0.23	.14	0.10	0.36	.79
Women	0.00	0.28	.99	−0.28	0.36	.43	0.03	0.58	.96
Men	−0.01	0.50	.98	−0.48	0.30	.11	0.19	0.45	.67
Diastolic BP									
All participants	0.08	0.18	.66	−0.34	0.16	.04	−0.02	0.26	.95
Women	−0.01	0.20	.97	−0.37	0.24	.12	−0.37	0.42	.38
Men	0.38	0.45	.40	−0.33	0.23	.15	0.31	0.34	.36

[a] BP: blood pressure.

[b] Models were adjusted for age, sex, family structure, reported antihypertensive drug use, and watch wear time.

[c] β represents the change in BP (mmHg) for every 1000 increase in steps.

Discussion

Principal Findings

In this study of community-dwelling participants, we measured habitual physical activity using a study smartwatch and home BP device. We observed an inverse association of higher habitual physical activity with lower home BP, even after adjusting for age, sex, watch wear time, antihypertensive drug use, and family structure. The association between physical activity and home BP, however, was rendered statistically nonsignificant after further adjusting for BMI, suggesting the mediating role of BMI.

When studied in community-based samples, the association of physical activity and BP remained controversial [3,20] and is perhaps dependent on the intensity of activity. For example, in a community-dwelling sample from the Dutch Lifelines cohort (n=125,402), higher self-reported commuting and leisure time moderate to vigorous physical activity was associated with lower BP in a dose-dependent manner (cross-sectional analysis) [1]. Similarly, in a sample of African American participants derived from the Jackson Heart Study (n=1311), lower baseline moderate to vigorous physical activity, but not overall physical activity, was associated with an increased risk of incident hypertension over a median follow-up of 8 years [2]. On the contrary, in 1717 participants of the Framingham Offspring study, the self-reported physical activity index (a composite score of daily physical activity and sedentary behavior) did not emerge as an independent predictor of incident hypertension in multivariable-adjusted models [6]. A similar lack of an independent association of physical activity with incident hypertension was observed in the participants of the National Health and Nutrition Examination Survey [5] and Coronary Artery Risk Development in (Young) Adults Study [4]. Since 2016, eFHS participants have undergone digital phenotyping using research smartwatches and BP cuffs, thereby enriching the existing information about these FHS participants [12]. With the goal of better understanding the relationship between physical activity and BP, in our current investigation, we studied the association of a common measure of physical activity (daily step count) [21] and home BP (a powerful predictor of adverse cardiovascular outcomes) [11]. The demographic-adjusted regression model (primary model) revealed an inverse association of habitual physical activity with home BP. However, with further adjustment for BMI (secondary model), the association became attenuated and nonsignificant, suggesting that BMI accounted for most or all of the association between habitual physical activity and home BP. The results of our exploratory BMI-stratified models further confirmed this observation as no significant association of habitual physical activity with home BP was observed in any BMI subgroups (normal weight, overweight, or obese).

Our findings, when taken together with prior observations, suggest that the effect of overall physical activity on BP in community-based settings is likely mediated via correlates of physical activity such as obesity. While several prior studies have reported an association of moderate to vigorous physical activity with BP and incident hypertension [1,2], we were unable to study the relationship of exercise intensity with home BP as the current wearables (including the Apple Watch used in our study) do not accurately discriminate between the intensity of exercise [22-24]. Nevertheless, community-dwelling individuals spend the majority of their time performing light physical activity (such as walking), and over the last decade, several studies have reported the beneficial effects of light physical activity on overall metabolic profile [25-27] and mortality [28]. It is plausible that light physical activity modulates BP via improvement in the metabolic profile (eg, BMI in our study) [29,30], whereas moderate to vigorous physical activity lowers BP more directly by reducing the vascular tone [3]. In fact, interventions aimed at increasing moderate to vigorous physical activity consistently lead to a reduction in BP in both normotensive and hypertensive individuals [7,8]. Of note, we observed an inverse association of home diastolic BP variability, an independent predictor of adverse cardiovascular outcomes [31], with higher step count independent of BMI. The association of habitual physical activity with reduced diastolic BP variability might be attributable to its effect on autonomic modulation [32]. Our findings suggest that overall habitual physical activity could lead to reduced cardiovascular risk despite the lack of an independent association with absolute home BP.

Another major implication of our study is that smartwatches provide scalable instruments to measure habitual physical activity in community-based settings. Prior studies have reported a significant reduction in BP-leveraging pedometer and smartphone-based physical activity promotion interventions [7]. Home BP is a stronger predictor of adverse cardiovascular outcomes compared with in-office BP [11], and self-monitoring of postexercise hypotension can improve exercise adherence [33]. Nearly one in five Americans currently own a smartwatch [34]. With the burgeoning field of wearable smart devices, our findings raise the possibility of leveraging smartwatches and home BP monitoring to promote physical activity to address the community burden of hypertension and associated comorbidities such as obesity [35].

Limitations

Our study has several limitations. First, the participants in our study were predominantly White and of European ancestry, and had a higher average daily step count (7595 steps/day) than the US average (4774 steps/day) [36]. Further, we acknowledge that our study sample (n=660) represents a subset of younger participants who attended the research center examination (n=3521). Although the baseline characteristics of the eFHS cohort and our study sample were similarly distributed, the possibility of selection bias exists. Our findings should be validated in samples with older participants and in racially diverse samples as there is evidence supporting the differential association of physical activity with incident hypertension, based on race [37]. Second, the analyses performed in our observational study are cross-sectional in nature. Therefore, our findings do not imply a causal association between lower PA and higher BP. However, our eFHS cohort is embedded in the overall FHS, and as these participants are systematically followed over the next few years, we would be able to study the pattern of habitual physical activity that relates to a higher risk of incident hypertension. Third, although the association between habitual physical activity and home BP was statistically significant, whether a 0.49 mmHg lower systolic BP and 0.36 mmHg lower diastolic BP per 1000 daily steps translates to a clinically meaningful impact in the community should be assessed in future studies. Even a 2 mmHg reduction in the average diastolic BP, however, would lead to a 17% decrease in hypertension prevalence [38]. From the clinical standpoint, considering the large number of daily steps required to achieve a modest reduction in BP, our findings might not translate to significant benefits at the individual level. Fourth, we relied on outputs from the Apple Watch proprietary algorithms to measure physical activity in our study. Although validated for measuring physical activity, these algorithms are not available publicly [22,39].

Conclusions

In this community-based sample of middle-aged adults who were enrolled in an electronic cohort at the time of their routine research center examination, we observed that higher habitual physical activity measured by a smartwatch was associated with a moderate, but statistically significant, reduction in home BP. Differences in BMI among study participants accounted for the majority of the observed association. The results of our study lay the ground work for leveraging smart devices to promote physical activity and improve the cardiometabolic phenotype in the community.

Acknowledgments

The Apple Watches were provided to Boston University by Apple, Inc at no cost to the study. This study was supported by an award from the Robert Wood Johnson Foundation (number 74624) and a grant from the National Heart Lung and Blood Institute (R01HL141434, HHSN268201500001I). Investigator time was supported from the following grants: R01HL126911 (EJB), 2R01 HL092577 (EJB), 2U54HL120163 (EJB), R01HL126911 (DDM), R01HL137734 (DDM), R01HL137794 (DDM), and R01HL13660 (DDM).

Conflicts of Interest

DDM discloses equity stakes or consulting relationships with Flexcon, Bristol-Myers Squibb, Boston Biomedical Associates, Pfizer, and Samsung, and research support from Sanofi Aventis, Otsuka Pharmaceuticals, Philips Healthcare, Biotronik, Bristol Myers Squibb, Boeringher Ingelheim, and Pfizer. DDM has an inventor stake in Mobile Sense Technologies, LLC. VK is a principal, and CN is an employee of CareEvolution, Inc, a health care technology company (https://careevolution.com). Apple was not involved in the study design, analysis, interpretation, or reporting of the study results. Starting 2020, EJB has been an uncompensated member for MyHeartLab Steering Committee, a primary investigator-initiated study from Samsung in the University of California San Francisco (PI: Jeffrey Olgin, MD). NS received funding from Novo Nordisk for an investigator-initiated research grant unrelated to the current paper. Other authors have no relevant disclosures.

Multimedia Appendix 1
Characteristics of the study participants compared to the overall electronic Framingham Heart Study cohort and all research center attendees during the study enrollment period.
[DOCX File , 15 KB - jmir_v23i6e25591_app1.docx]

Multimedia Appendix 2
Additional methods and results.
[DOCX File , 60 KB - jmir_v23i6e25591_app2.docx]

Multimedia Appendix 3
Association of log-transformed daily step count with home blood pressure using separate mixed linear effect models for systolic and diastolic blood pressure.
[DOCX File , 15 KB - jmir_v23i6e25591_app3.docx]

Multimedia Appendix 4
Association of daily step count with home blood pressure in participants with 60 or more active days.
[DOCX File , 15 KB - jmir_v23i6e25591_app4.docx]

Multimedia Appendix 5
Association of daily step count with home blood pressure in participants with 90 or more active days.
[DOCX File , 15 KB - jmir_v23i6e25591_app5.docx]

Multimedia Appendix 6

Association of daily step count with home blood pressure in participants with nine or more blood pressure readings.
[DOCX File , 15 KB - jmir_v23i6e25591_app6.docx]

Multimedia Appendix 7
Association of daily step count with home blood pressure using a threshold of 10 hours/day to define active days.
[DOCX File , 15 KB - jmir_v23i6e25591_app7.docx]

Multimedia Appendix 8
Box plots depicting average daily step count stratified by BMI categories. N=200 for normal weight, N=277 for overweight, and N=183 for obese.
[PNG File , 11 KB - jmir_v23i6e25591_app8.png]

Multimedia Appendix 9
Box plots depicting average home systolic and diastolic blood pressure stratified by BMI categories. N=200 for normal weight, N=277 for overweight, and N=183 for obese. DBP: diastolic blood pressure; SBP: systolic blood pressure.
[PNG File , 20 KB - jmir_v23i6e25591_app9.png]

Multimedia Appendix 10
Scatter plots depicting the correlation between average daily step count and home blood pressure stratified by BMI categories.
[PNG File , 45 KB - jmir_v23i6e25591_app10.png]

Multimedia Appendix 11
Baseline characteristics of participants with and without a history of hypertension.
[DOCX File , 14 KB - jmir_v23i6e25591_app11.docx]

Multimedia Appendix 12
Association of daily step count with home blood pressure in participants with a history of hypertension.
[DOCX File , 14 KB - jmir_v23i6e25591_app12.docx]

Multimedia Appendix 13
Association of daily step count with home blood pressure tertile.
[DOCX File , 15 KB - jmir_v23i6e25591_app13.docx]

References

1. Byambasukh O, Snieder H, Corpeleijn E. Relation Between Leisure Time, Commuting, and Occupational Physical Activity With Blood Pressure in 125 402 Adults: The Lifelines Cohort. J Am Heart Assoc 2020 Feb 18;9(4):e014313 [FREE Full text] [doi: 10.1161/JAHA.119.014313] [Medline: 32067583]
2. Diaz KM, Booth JN, Seals SR, Abdalla M, Dubbert PM, Sims M, et al. Physical Activity and Incident Hypertension in African Americans: The Jackson Heart Study. Hypertension 2017 Mar;69(3):421-427 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.116.08398] [Medline: 28137988]
3. Diaz KM, Shimbo D. Physical activity and the prevention of hypertension. Curr Hypertens Rep 2013 Dec;15(6):659-668 [FREE Full text] [doi: 10.1007/s11906-013-0386-8] [Medline: 24052212]
4. Dyer AR, Liu K, Walsh M, Kiefe C, Jacobs DR, Bild DE. Ten-year incidence of elevated blood pressure and its predictors: the CARDIA study. Coronary Artery Risk Development in (Young) Adults. J Hum Hypertens 1999 Jan;13(1):13-21 [FREE Full text] [doi: 10.1038/sj.jhh.1000740] [Medline: 9928747]
5. Ford ES, Cooper RS. Risk factors for hypertension in a national cohort study. Hypertension 1991 Nov;18(5):598-606. [doi: 10.1161/01.hyp.18.5.598] [Medline: 1937662]
6. Parikh NI, Pencina MJ, Wang TJ, Benjamin EJ, Lanier KJ, Levy D, et al. A risk score for predicting near-term incidence of hypertension: the Framingham Heart Study. Ann Intern Med 2008 Jan 15;148(2):102-110. [doi: 10.7326/0003-4819-148-2-200801150-00005] [Medline: 18195335]
7. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
8. Cornelissen VA, Smart NA. Exercise training for blood pressure: a systematic review and meta-analysis. J Am Heart Assoc 2013 Feb 01;2(1):e004473 [FREE Full text] [doi: 10.1161/JAHA.112.004473] [Medline: 23525435]
9. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003 Jun;37(3):197-206; discussion 206 [FREE Full text] [doi: 10.1136/bjsm.37.3.197] [Medline: 12782543]

10. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
11. Bobrie G, Chatellier G, Genes N, Clerson P, Vaur L, Vaisse B, et al. Cardiovascular prognosis of "masked hypertension" detected by blood pressure self-measurement in elderly treated hypertensive patients. JAMA 2004 Mar 17;291(11):1342-1349. [doi: 10.1001/jama.291.11.1342] [Medline: 15026401]
12. McManus DD, Trinquart L, Benjamin EJ, Manders ES, Fusco K, Jung LS, et al. Design and Preliminary Findings From a New Electronic Cohort Embedded in the Framingham Heart Study. J Med Internet Res 2019 Mar 01;21(3):e12143 [FREE Full text] [doi: 10.2196/12143] [Medline: 30821691]
13. Schnabel RB, Larson MG, Yamamoto JF, Sullivan LM, Pencina MJ, Meigs JB, et al. Relations of biomarkers of distinct pathophysiological pathways and atrial fibrillation incidence in the community. Circulation 2010 Jan 19;121(2):200-207 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.882241] [Medline: 20048208]
14. Goldberg EM, Levy PD. New Approaches to Evaluating and Monitoring Blood Pressure. Curr Hypertens Rep 2016 Jun;18(6):49 [FREE Full text] [doi: 10.1007/s11906-016-0650-9] [Medline: 27137524]
15. Topouchian J, Agnoletti D, Blacher J, Youssef A, Chahine MN, Ibanez I, et al. Validation of four devices: Omron M6 Comfort, Omron HEM-7420, Withings BP-800, and Polygreen KP-7670 for home blood pressure measurement according to the European Society of Hypertension International Protocol. Vasc Health Risk Manag 2014;10:33-44 [FREE Full text] [doi: 10.2147/VHRM.S53968] [Medline: 24476688]
16. 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]
17. Thomas L, Williams M. Promoting physical activity in the workplace: using pedometers to increase daily activity levels. Health Promot J Austr 2006 Aug;17(2):97-102. [doi: 10.1071/he06097] [Medline: 16916311]
18. Hond ED, Celis H, Fagard R, Keary L, Leeman M, O'Brien E, THOP investigators. Self-measured versus ambulatory blood pressure in the diagnosis of hypertension. J Hypertens 2003 Apr;21(4):717-722. [doi: 10.1097/00004872-200304000-00014] [Medline: 12658017]
19. Celis H, De Cort P, Fagard R, Thijs L, Staessen JA. For how many days should blood pressure be measured at home in older patients before steady levels are obtained? J Hum Hypertens 1997 Oct;11(10):673-677 [FREE Full text] [doi: 10.1038/sj.jhh.1000517] [Medline: 9400910]
20. Igarashi Y, Akazawa N, Maeda S. The required step count for a reduction in blood pressure: a systematic review and meta-analysis. J Hum Hypertens 2018 Dec;32(12):814-824 [FREE Full text] [doi: 10.1038/s41371-018-0100-z] [Medline: 30127487]
21. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of Step Volume and Intensity With All-Cause Mortality in Older Women. JAMA Intern Med 2019 Aug 01;179(8):1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
22. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
23. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
24. Falter M, Budts W, Goetschalckx K, Cornelissen V, Buys R. Accuracy of Apple Watch Measurements for Heart Rate and Energy Expenditure in Patients With Cardiovascular Disease: Cross-Sectional Study. JMIR Mhealth Uhealth 2019 Mar 19;7(3):e11889 [FREE Full text] [doi: 10.2196/11889] [Medline: 30888332]
25. Whitaker KM, Pettee Gabriel K, Buman MP, Pereira MA, Jacobs DR, Reis JP, et al. Associations of Accelerometer-Measured Sedentary Time and Physical Activity With Prospectively Assessed Cardiometabolic Risk Factors: The CARDIA Study. J Am Heart Assoc 2019 Jan 08;8(1):e010212 [FREE Full text] [doi: 10.1161/JAHA.118.010212] [Medline: 30616480]
26. Healy GN, Dunstan DW, Salmon J, Cerin E, Shaw JE, Zimmet PZ, et al. Objectively measured light-intensity physical activity is independently associated with 2-h plasma glucose. Diabetes Care 2007 Jun;30(6):1384-1389. [doi: 10.2337/dc07-0114] [Medline: 17473059]
27. Green AN, McGrath R, Martinez V, Taylor K, Paul DR, Vella CA. Associations of objectively measured sedentary behavior, light activity, and markers of cardiometabolic health in young women. Eur J Appl Physiol 2014 May;114(5):907-919. [doi: 10.1007/s00421-014-2822-0] [Medline: 24463602]
28. LaMonte MJ, Buchner DM, Rillamas-Sun E, Di C, Evenson KR, Bellettiere J, et al. Accelerometer-Measured Physical Activity and Mortality in Women Aged 63 to 99. J Am Geriatr Soc 2018 May;66(5):886-894 [FREE Full text] [doi: 10.1111/jgs.15201] [Medline: 29143320]
29. Iwane M, Arita M, Tomimoto S, Satani O, Matsumoto M, Miyashita K, et al. Walking 10,000 steps/day or more reduces blood pressure and sympathetic nerve activity in mild essential hypertension. Hypertens Res 2000 Nov;23(6):573-580 [FREE Full text] [doi: 10.1291/hypres.23.573] [Medline: 11131268]

30. Gerage AM, Benedetti TRB, Farah BQ, Santana FDS, Ohara D, Andersen LB, et al. Sedentary Behavior and Light Physical Activity Are Associated with Brachial and Central Blood Pressure in Hypertensive Patients. PLoS One 2015;10(12):e0146078 [FREE Full text] [doi: 10.1371/journal.pone.0146078] [Medline: 26717310]
31. Johansson JK, Niiranen TJ, Puukka PJ, Jula AM. Prognostic value of the variability in home-measured blood pressure and heart rate: the Finn-Home Study. Hypertension 2012 Feb;59(2):212-218. [doi: 10.1161/HYPERTENSIONAHA.111.178657] [Medline: 22215704]
32. Tebar WR, Ritti-Dias RM, Mota J, Farah BQ, Saraiva BTC, Damato TMM, et al. Relationship between domains of physical activity and cardiac autonomic modulation in adults: a cross-sectional study. Sci Rep 2020 Sep 23;10(1):15510 [FREE Full text] [doi: 10.1038/s41598-020-72663-7] [Medline: 32968194]
33. Zaleski AL, Taylor BA, Park CL, Santos LP, Panza G, Kramarz M, et al. Using the immediate blood pressure benefits of exercise to improve exercise adherence among adults with hypertension: a randomized clinical trial. J Hypertens 2019 Sep;37(9):1877-1888. [doi: 10.1097/HJH.0000000000002115] [Medline: 31058797]
34. Mobile Fact Sheet. Pew Research Center. 2021. URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2021-05-15]
35. Haufe S, Kerling A, Protte G, Bayerle P, Stenner HT, Rolff S, et al. Telemonitoring-supported exercise training, metabolic syndrome severity, and work ability in company employees: a randomised controlled trial. The Lancet Public Health 2019 Jul;4(7):e343-e352. [doi: 10.1016/S2468-2667(19)30075-1] [Medline: 31204284]
36. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Jul 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
37. Duncan DT, Quarells RC, Din-Dzietham R, Arroyo C, Davis SK. Physical activity and incident hypertension among blacks: no relationship? Prev Chronic Dis 2006 Jul;3(3):A109 [FREE Full text] [Medline: 16776870]
38. Whelton PK, He J, Appel LJ, Cutler JA, Havas S, Kotchen TA, National High Blood Pressure Education Program Coordinating Committee. Primary prevention of hypertension: clinical and public health advisory from The National High Blood Pressure Education Program. JAMA 2002 Oct 16;288(15):1882-1888. [doi: 10.1001/jama.288.15.1882] [Medline: 12377087]
39. Veerabhadrappa P, Moran MD, Renninger MD, Rhudy MB, Dreisbach SB, Gift KM. Tracking Steps on Apple Watch at Different Walking Speeds. J Gen Intern Med 2018 Jun;33(6):795-796 [FREE Full text] [doi: 10.1007/s11606-018-4332-y] [Medline: 29633143]

Abbreviations

BP: blood pressure
eFHS: electronic Framingham Heart Study
FHS: Framingham Heart Study

Edited by R Kukafka; submitted 07.11.20; peer-reviewed by M Heneghan, B Sternfeld, M Falter; comments to author 22.12.20; revised version received 22.02.21; accepted 16.03.21; published 24.06.21.

Please cite as:
Sardana M, Lin H, Zhang Y, Liu C, Trinquart L, Benjamin EJ, Manders ES, Fusco K, Kornej J, Hammond MM, Spartano N, Pathiravasan CH, Kheterpal V, Nowak C, Borrelli B, Murabito JM, McManus DD
Association of Habitual Physical Activity With Home Blood Pressure in the Electronic Framingham Heart Study (eFHS): Cross-sectional Study
J Med Internet Res 2021;23(6):e25591
URL: https://www.jmir.org/2021/6/e25591/
doi:10.2196/25591
PMID:34185019

Original Paper

Evaluation Framework for Successful Artificial Intelligence–Enabled Clinical Decision Support Systems: Mixed Methods Study

Mengting Ji[1,2], MD; Georgi Z Genchev[3,4,5], PhD; Hengye Huang[1], MD; Ting Xu[1], BA; Hui Lu[3,4,6], PhD; Guangjun Yu[1,7], PhD

[1]School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2]Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China
[3]Center for Biomedical Informatics, Shanghai Children's Hospital, Shanghai, China
[4]SJTU-Yale Joint Center for Biostatistics, Shanghai Jiao Tong University, Shanghai, China
[5]Bulgarian Institute for Genomics and Precision Medicine, Sofia, Bulgaria
[6]Department of Bioinformatics and Biostatistics, Shanghai Jiao Tong University, Shanghai, China
[7]Shanghai Children’s Hospital, Shanghai Jiao Tong University, Shanghai, China

Corresponding Author:
Guangjun Yu, PhD
Shanghai Children’s Hospital
Shanghai Jiao Tong University
No 355 Luding Road
Shanghai, 200062
China
Phone: 86 18917762998
Email: yugj1688@shchildren.com.cn

Abstract

Background: Clinical decision support systems are designed to utilize medical data, knowledge, and analysis engines and to generate patient-specific assessments or recommendations to health professionals in order to assist decision making. Artificial intelligence–enabled clinical decision support systems aid the decision-making process through an intelligent component. Well-defined evaluation methods are essential to ensure the seamless integration and contribution of these systems to clinical practice.

Objective: The purpose of this study was to develop and validate a measurement instrument and test the interrelationships of evaluation variables for an artificial intelligence–enabled clinical decision support system evaluation framework.

Methods: An artificial intelligence–enabled clinical decision support system evaluation framework consisting of 6 variables was developed. A Delphi process was conducted to develop the measurement instrument items. Cognitive interviews and pretesting were performed to refine the questions. Web-based survey response data were analyzed to remove irrelevant questions from the measurement instrument, to test dimensional structure, and to assess reliability and validity. The interrelationships of relevant variables were tested and verified using path analysis, and a 28-item measurement instrument was developed. Measurement instrument survey responses were collected from 156 respondents.

Results: The Cronbach α of the measurement instrument was 0.963, and its content validity was 0.943. Values of average variance extracted ranged from 0.582 to 0.756, and values of the heterotrait-monotrait ratio ranged from 0.376 to 0.896. The final model had a good fit (χ_{26}^2=36.984; *P*=.08; comparative fit index 0.991; goodness-of-fit index 0.957; root mean square error of approximation 0.052; standardized root mean square residual 0.028). Variables in the final model accounted for 89% of the variance in the user acceptance dimension.

Conclusions: User acceptance is the central dimension of artificial intelligence–enabled clinical decision support system success. Acceptance was directly influenced by perceived ease of use, information quality, service quality, and perceived benefit. Acceptance was also indirectly influenced by system quality and information quality through perceived ease of use. User acceptance and perceived benefit were interrelated.

(J Med Internet Res 2021;23(6):e25929) doi:10.2196/25929

KEYWORDS

artificial intelligence; AI; clinical decision support systems; evaluation framework

Introduction

Clinical Decision Support Systems

Clinical decision support systems are computer-based enterprise systems designed to utilize massive data, medical knowledge, and analysis engines as well as to generate patient-specific assessments or recommendations to health professionals in order to assist clinical decision making through human–computer interaction [1,2]. These systems provide services ranging from simple reminders to complex risk prediction [3] and support health care providers in diagnosis, treatment decisions, and population health management. Clinical decision support systems assist one or more levels of decision making: alerting, interpreting, critiquing, assisting, diagnosing, and managing [4]. Diagnostic support systems are a subset of clinical decision support systems that are specifically designed to support clinician in diagnosing patients [5]. Artificial intelligence (AI)–enabled clinical decision support systems combine the knowledge reasoning techniques of AI and the functional models of clinical decision support systems [6].

AI-Enabled Clinical Decision Support Systems: Characteristics, Usage, and Benefits

AI-enabled clinical decision support systems include an intelligent component [6], and in comparison to traditional clinical decision support systems, represent a paradigm shift. They are designed to aid clinicians by converting raw medical-related data, documents, and expert practice into a set of sophisticated algorithms, applying techniques such as machine learning, knowledge graphs, natural language processing, and computer vision so that users find suitable solutions to their medical problems and make clinical decisions [7]. AI-enabled clinical decision support systems have the potential to improve clinicians' performance, quality of health care, and patient safety [8].

Diagnostics are a primary use case of AI-enabled clinical decision support systems, and these systems have been applied in the field of rare disease diagnosis [9], sepsis detection or prediction [10], fracture detection [11], and cancer detection or diagnosis [12,13]. In addition, current AI-enabled clinical decision support systems are also used in medication therapy [14,15] and health care management [16,17].

The greatest benefits of AI-enabled clinical decision support systems reside in their ability to learn from real-world use and experience (ie, training) and their capabilities for improving their performance (ie, adaptation) [18]. By using techniques such as knowledge graphs and natural language processing, AI can deal with large amounts of text classification, information retrieval, and information extraction from the corpora that is provided by hospital electronic health records. Based on structured data, AI can support more comprehensive and more personalized decision-making suggestions for clinicians through techniques such as machine learning. Another benefit is that the functionality and utility from combining clinical decision support systems with AI techniques surpass those of traditional clinical decision support systems, and the system improves and supports the decision-making process by providing intelligent behavioral patterns, with the ability to learn new clinical knowledge [7].

Need for AI-Enabled Clinical Decision Support System Evaluation

A comprehensive evaluation framework with common elements and interoperability is necessary to serve as a reference for AI-enabled clinical decision support system design and evaluation, with focuses on cross-disciplinary communication and collaboration, and there is a pressing need to develop robust methodologies and empirically based tools for such evaluation. The factors driving this need are the uncertain added value of AI-enabled clinical decision support system implementation, lack of attention, and the possible benefits of comprehensive evaluation implementations.

First, the added value of AI-enabled clinical decision support system implementations in a clinical setting is not firmly established, though evidence exists that such implementations offer potential benefit to patients, clinicians, and health care in general [19]. Introducing this type of system in clinical settings is not without risk [8]. Similar to any other newly introduced technology, AI-enabled clinical decision support systems may disrupt clinical service, threaten patient safety [20], and cause more negative than positive impacts [19]. As a result, there are concerns that AI-enabled clinical decision support system implementation can introduce new errors and have unintended consequences [21]. Additionally, the effect of these systems on clinical, social, and economic outcomes is still controversial which highlights the need to evaluate recognized value parameters [22]. Second, attention to evaluation of clinical decision support systems, in general, and AI-enabled clinical decision support systems, in particular, remains weak [23], which has resulted in a paucity of data on safety, effectiveness, cost benefits, and impacts of AI-enabled clinical decision support systems on patients and health systems [24,25]. Finally, the evaluation of AI-enabled clinical decision support systems is a learning and knowledge-gaining process, and it also helps to identify the gaps to be filled [26]. Findings of comprehensive evaluations could be used to help improve implementations [27].

AI-Enabled Clinical Decision Support System Evaluation Methodologies

The approach to AI-enabled clinical decision support system evaluation is influenced by a sociotechnical regime, which informs and guides the development of the robust and focused evaluation method of this study. It has increasingly been acknowledged that evaluations of such systems are based on a sociological understanding of the complex practices in which the information technologies are to function [28]. A careful balance between social and technical value is required in order to ensure that unwanted consequences do not pose a threat to patients [26] and clinical practices.

A well-defined success measure, based on users' perspectives, that specifies aspects of AI-enabled clinical decision support systems that determine their success [29] is critical for a robust performance and usefulness evaluation framework. Due to the user-centric nature of information system development and evaluation [30,31], evaluation of AI-enabled clinical decision support system success aims to recognize factors relevant to user acceptance and utility, thus analysis of articulated users' opinions is necessary [32]. Clinicians are the direct users of AI-enabled clinical decision support systems; the adoption of the product depends on the individual physicians who decide to use it [5]. In many scenarios, clinicians make decisions for patients, and clinicians are responsible for the medical decisions they make. Predicting and managing users' attitudes toward AI-enabled clinical decision support systems lead to an in-depth understanding of these systems via situated practice [33] and help developers and medical managers maximize user acceptance. Lack of a well-defined success measure is likely to lead to inappropriate evaluation that does not reflect the clinical impact of AI-enabled clinical decision support systems and may hamper technology advancement[19].

A comprehensive evaluation methodology involves a multidisciplinary process and diverse stakeholder involvement, which, when applied to AI-enabled clinical decision support system evaluation, refers to a mixed methodology not only based on tenets in medicine and information technology but also social and cognitive psychology [30]. Using both qualitative and quantitative methods within a single research project has been shown to provide a richer understanding of a given topic than using solely either a qualitative or quantitative approach, facilitate better and more accurate inferences, and provide an integrated perspective [34]. A similar benefit would likely apply when employing mixed methods in designing an AI-enabled clinical decision support system evaluation scheme.

AI-enabled clinical decision support system interface with a diverse set of clinical and nonclinical users and stakeholders whose inputs are integral to the evaluation process. Health care enterprises are multiprofessional organizations that often include dual hierarchical structures involving clinical practitioners and managers [35], and in such settings, AI-enabled clinical decision support systems are not only tools for clinical practitioners who interact directly with the system (eg, physicians, nurses, pharmacists) but also for nonclinical workers (eg, medical administrators). Additionally, there is still an important group of invisible stakeholders, namely patients, who can be affected by these systems use even without direct interaction. The relationships of such diverse groups of stakeholders can prove to be complex, with competing interests and values; therefore, the views, beliefs, and assumptions of stakeholders must be exposed and considered within the AI-enabled clinical decision support system evaluation process [33,36].

Objective

We aimed to address the gap in evaluation knowledge and methodologies by identifying which variables influence AI-enabled clinical decision support system success and using these variables to develop a parsimonious evaluation framework. Specifically, we (1) proposed an evaluation framework with 6 variables and hypotheses about interrelationships between the 6 variables based on the literature review, (2) developed and validated an instrument using the 6 variables for assessing the success of diagnostic AI-enabled clinical decision support systems, and (3) tested the hypotheses using path analysis with latent variables in a structural equation model.

Methods

Ethics Approval

This study was approved by the Ethics Review Committee, Children's Hospital of Shanghai/Shanghai Children's Hospital, Shanghai Jiao Tong University (file number 2020R050-E01).

Overview

Our study combined qualitative and quantitative methodologies to validate a proposed evaluation framework, which consisted of a model with hypotheses and containing 6 variables.. A Chinese-language measurement instrument was developed with the goal to measure and quantify the 6 variables, following established instrument development paradigm. A literature review and a Delphi process were conducted to develop the measurement instrument items, cognitive interviews, pretest, and web-based survey. Exploratory factor analysis was used to construct the constituent questions of the measurement instrument, reliability and validity tests were performed, and the interrelations of the variables were tested and verified.

Theory

Evaluation methodologies are informed by a rich corpus of theory, which provides a robust foundation for designing an AI-enabled clinical decision support system evaluation framework. In this study and in previous review work [37], three classic theories were used, namely, the DeLone and McLean Model of Information Systems Success [38], the Information Systems Continuance Model [39,40], and the Information Value Chain Theory [29].

An updated model of information systems success that captures multidimensionality and interdependency was proposed by DeLone and McLean in 2003 [38]; the model is a basic and flexible framework of information system evaluation that can adapt to the complexity of the clinical environment [41-44]. In considering the importance of user acceptance and retention to an information system's success, the information systems continuance model describes the path from expectation confirmation to the formation of users' intention to continuance [39]. The information value chain theory underlines decision improvement as the main purpose of technology and provides a mechanism to separate process outcomes from clinical outcomes [45].

Evaluation Framework Model Variable and Measurement Instrument Item Selection

Literature Search

A set of evaluation model variables and a candidate set of medical AI and clinical decision support system evaluation items were collected through a literature review [35]. A broad search strategy was employed, using multiple databases including Cochrane, MEDLINE, EMBASE, Web of Science, PubMed, CINAHL, PsycINFO, and INSPEC. Studies published from January 2009 to May 2020 were utilized to inform the clinical decision support system evaluation items selection and studies published January 2009 to April 2020 for the AI evaluation items discovery. A candidate set of 6 model variables (Multimedia Appendix 1) and a candidate set of 45 evaluation items were identified.

Delphi Process

The candidate set of evaluation items was examined and finalized using a Delphi process. Delphi is a structured group communication process, designed to obtain a consensus of opinion from a group of experts [46].

Snowball sampling was used to identify a group of experts. Expert selection criteria were (1) clinical practitioners who worked in a medical specialty at least 10 years, preferably had a PhD (minimum postgraduate qualification), had a professional title at the advanced level or above, had an appointment or affiliation with a professional organization, and had more than 1 year of practical experience (with respect to AI-enabled clinical decision support systems); (2) hospital chief information officers who worked in an information system specialty at least 10 years, had a postgraduate qualification, had a midlevel professional title or above, and had an appointment or affiliation with a professional information system organization; or (3) information technology engineers working in medical information system enterprises who worked in AI or clinical decision support systems at least 5 years, had a postgraduate qualification, and had a midlevel position title or above.

In addition to these selection criteria, a measure of degree of expert authority was introduced to add or remove experts from each round of the Delphi process. The degree of expert authority C_r was defined $C_r = (C_a + C_s) / 2$, using 2 self-evaluated scores—C_a is their familiarity with the problem, and C_s is their knowledge base to judge the program. C_s and C_a ranged between 1 and 5, with a higher value indicating more reliable judgment and more familiarity with the problem. If the self-rated degree of expert authority was >3, the expert was retained, otherwise the expert was removed from group. As a result, a total of 11 experts were selected from diverse areas of expertise and professional focus: clinical practitioners, hospital chief information officers, and information technology engineers working in medical information system enterprises.

The experts were invited to participate in the modified Delphi process via email. Those who accepted were sent an email with a link to the round 1 consultation. Experts were required to provide a relevance score for each item in the candidate set using a 4-point Likert scale (1=not relevant, 2=relevant but requires major revision, 3=relevant but requires minor revision, 4=very relevant and requires no revision). Experts were given 2 weeks to complete each round. A reminder was sent 2 days before the deadline to those who had not completed the survey. The 2-round Delphi process was carried out from May to July 2020.

The content validity was assessed in the last round of the Delphi process. Item-content validity was calculated as the percentage of expert ratings ≥3; if item-content validity was ≥0.8 (ie, expert endorsement), the item was retained. The mean item-content validity, representing the content validity of the measurement instrument of all retained items from the last round was computed. At the end of this step, the set of evaluation items for the measurement instrument were finalized. The final set consisted of 29 evaluation items.

Measurement Instrument Refinement

The measurement instrument consisted of the set of evaluation items measured by a web-based survey. A draft set of survey questions was refined by employing cognitive interviews and a pretest. Interviewees (n=5) who were postgraduates majoring in health informatics or end-users of AI-enabled clinical decision support systems (ie, clinicians) were asked to verbalize the mental process entailed in providing answers. The pretest included 20 end-users. The interviews and pretest were conducted in July 2020 and aimed to assess the extent to which the survey questions reflected the domain of interest and that answers produced valid measurements. Responses used a Likert scale from 1 (strongly disagree) to 7 (strongly agree). The wording of the questions was subsequently modified based on the feedback from the respondents. The web-based survey was initiated in July and was closed in September 2020.

Study Population

The evaluation entities chosen in this study were AI-enabled clinical decision support systems designed to support the risk assessment of venous thromboembolism among inpatients: AI-enabled clinical decision support systems that automatically capture electronic medical records based on natural language processing supporting assessment based on individual risk of thrombosis (eg, Caprini scale or Wells scoring), with monitoring of users and reminders sent to users to provide additional data were targeted.

Survey Participants and Sample Size

Users of target AI-enabled clinical decision support systems who had at least 1 month of user experience were included. The convenience sample participants were based in 3 hospitals in Shanghai that implemented venous thromboembolism risk assessment AI-enabled clinical decision support systems in clinical settings. We appointed an investigator at each hospital site who was responsible for stating the objective of the study, for identifying target respondents, and for monitoring the length of time it took the participants to complete the survey. This was a voluntary survey. The investigators transmitted the electronic questionnaire link to the respondents through the WeChat communication app.

To ensure usability for exploratory factor analysis [47] and to obtain parameter estimates with standard errors small enough to be of practical use in structural equation modeling [48,49], the required sample size was calculated using to participant-to-item ratio (ranging from 5:1 to 10:1), yielding n=150. A response rate ≥70% was targeted to support external validity [50].

Quality Control Measures

Quality control measures were implemented to ensure logical consistency, with completeness checks before the questionnaire was submitted by the responders. Before submitting, respondents could review or change their answers. In order to avoid duplicates caused by repeat submissions, respondents accessed the survey via a WeChat account. Submitted questionnaires meeting the following criteria were deleted: (1) filling time <100 seconds, or (2) the answer of following 2 questions were contradictory: “How often do you use the AI-enabled clinical decision support systems?” versus “You use the AI-enabled clinical decision support systems frequently.” Finally, we asked the point-of-contact individuals in each hospital to send online notifications to survey respondents at least 3 times at regular intervals in order to improve the response rate.

Statistical Analysis

Overview

Statistical analyses were performed (SPSS Amos, version 21, IBM Corp) to (1) identify items of measurement instrument that were not related to AI-enabled clinical decision support system success for deletion, (2) explore the latent constructs of the measurement instrument, and (3) evaluate reliability and validity of the measurement instrument.

Measurement Instrument Item Reduction

Critical ratio and significance were calculated using independent *t* tests between high- (upper 27%) and low- (lower 27%) score groups. Item-scale correlation was calculated using Pearson correlation. Corrected item-to-total correlations and the effect on Cronbach α if an item was deleted were calculated using reliability analysis. Item-scale correlation and corrected item-to-total correlations were indications of the degree to which each item was correlated with the total score. Criteria for potential elimination were (1) nonsignificant critical ratio (*P*>.05), (2) item-scale correlation <0.40, (3) corrected item-to-total correlation <0.40, (4) an increased α if the item was deleted [51,52], that is, if α increased with an item removed, we considered removal of the item from the measurement instrument [49].

Latent Construct of Measurement Instrument

Construct of the measurement tool was tested using exploratory factor analysis. Principal component analysis was applied for factor extraction, and the Promax with Kaiser normalization rotation strategy was used to redefine the factors to improve their interpretability. The cutoff strategy was based on verify if the data set was suitable for exploratory factor analysis—the Bartlett test of sphericity should be statistically significant (*P*<.05) and a Kaiser-Meyer-Olkin value ≥.60 is considered mediocre [49], a value ≥.90 is marvelous [53]. Only factors with an eigenvalue ≥0.50 were retained.

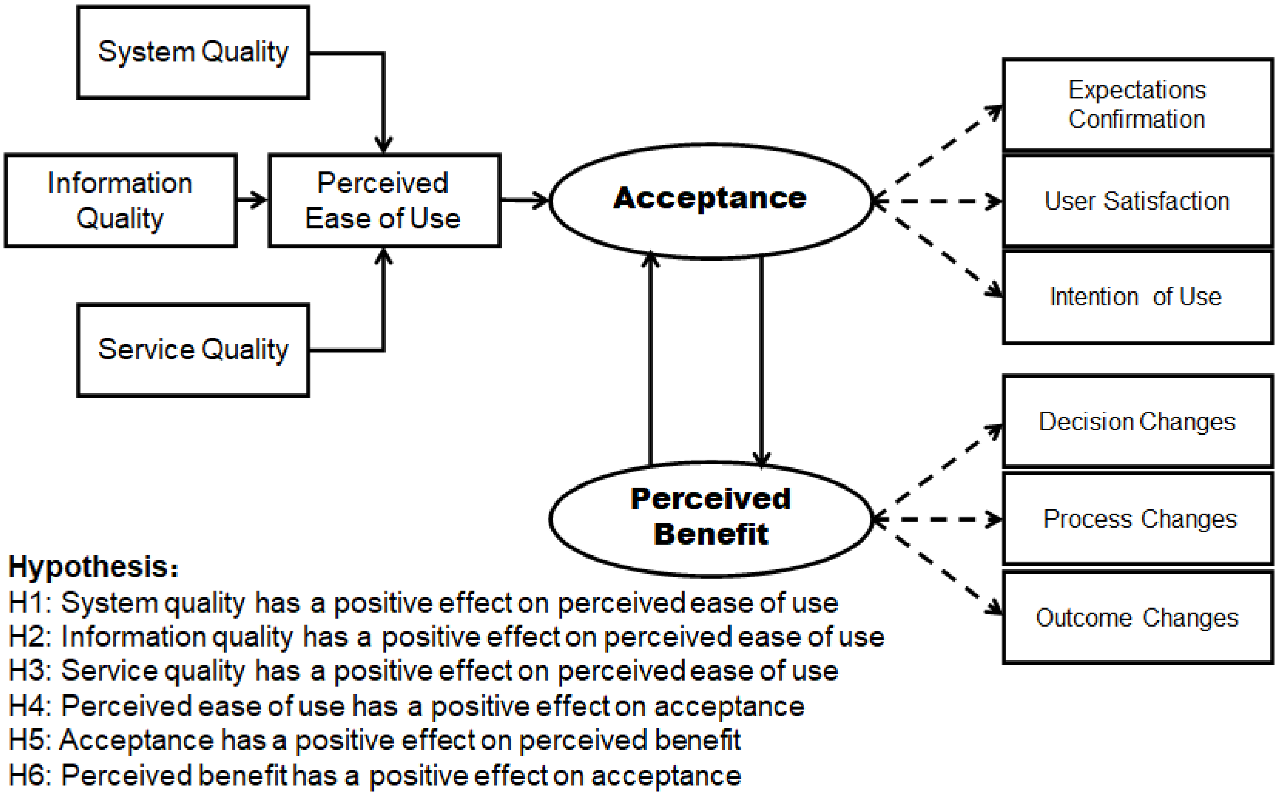
Reliability and Validity of Measurement Instrument

Cronbach α coefficients were calculated to assess internal consistencies of the scale and each subscale; values >.80 are preferred [49,50]. Convergent validity and discriminant validity were tested using maximum likelihood estimation confirmatory factor analysis in structural equation modeling. Average variance extracted was used as an indicator of convergent validity, and values >.50 were considered acceptable. The heterotrait-monotrait ratio of correlations was used to test discriminant validity. A heterotrait-monotrait ratio value <0.90 provided sufficient evidence of the discriminant validity of constructs [54].

Path Analysis

Interrelationships between variables selected for the evaluation framework were hypothesized in a model (Figure 1). The model was tested using path analysis with latent variables in structural equation modeling. We used the following indicators to assess competence of the model fit: chi-square (significant if *P*>.05), ratio of chi-square to degrees of freedom <2.00), comparative fit index >0.95, goodness-of-fit index >0.95, root mean square error of approximation <0.06, and standardized root mean square residual ≤0.08 [52,55].

Figure 1. Evaluation model hypotheses.



Results

Measurement Instrument

Delphi Process and Evaluation Item Selection

Of the 11 experts invited to participate (Multimedia Appendix 2), all accepted in round 1 (100% response rate) and 10 accepted in round 2 (91% response rate). Most respondents in round 2 (9/10, 90%) identified themselves as expert or very expert ($C_r \geq 4.0$) with respect to AI-enabled clinical decision support systems. Consensus was reached in round 2: 29 items obtained at least 80% endorsement (Table 1).

Table 1. Accepted items in the Delphi process.

Variables and items	Item-content validity	Critical ratio[a] (*t* value)	Item-scale correlation[a]	Corrected item-to-total correlation	Cronbach α if item was deleted
Perceived ease of use					
Learnability	1.00	6.419	0.643	0.615	.961
Operability	1.00	7.384	0.628	0.596	.961
User interface	0.90	10.496	0.700	0.669	.960
Data entry	1.00	10.530	0.655	0.622	.961
Advice display	1.00	7.938	0.655	0.621	.961
Legibility	1.00	7.836	0.666	0.641	.961
System quality					
Response time	1.00	7.826	0.606	0.565	.961
Stability	1.00	7.949	0.541	0.498	.962
Information quality					
Security	1.00	9.247	0.588	0.560	.961
Diagnostic performance	1.00	11.346	0.746	0.726	.960
Decision changes					
Changes in order behavior	0.90	8.593	0.667	0.637	.961
Changes in diagnosis	0.90	8.843	0.634	0.600	.961
Process changes					
Productivity	1.00	11.112	0.726	0.699	.960
Effectiveness	1.00	14.078	0.840	0.823	.959
Overall usefulness	1.00	13.720	0.826	0.809	.959
Adherence to standards	1.00	8.843	0.711	0.688	.960
Medical quality	1.00	8.945	0.717	0.696	.960
User knowledge and skills	0.80	8.366	0.715	0.692	.960
Outcome changes					
Change in clinical outcomes	0.90	10.974	0.741	0.719	.960
Change in patient-reported outcomes	0.80	10.769	0.716	0.692	.960
Service quality					
Operation and maintenance	0.90	9.624	0.590	0.555	.961
Information updating to keep timeliness	1.00	9.601	0.640	0.614	.961
Acceptance					
Usage	0.80	4.686	0.323[b]	0.282[b]	.963[b]
Expectations confirmation	1.00	14.174	0.856	0.841	.959
Satisfaction of system quality	0.80	12.248	0.816	0.798	.959
Satisfaction of information quality	0.80	13.437	0.828	0.813	.959
Satisfaction of service quality	0.80	11.031	0.737	0.714	.960
Overall satisfaction	1.00	15.053	0.873	0.860	.959
Intention of use	0.90	13.500	0.855	0.840	.959

[a] For all values in this column, $P<.001$.

[b] Based on this value, the item meets the standard for potential deletion.

Measurement Instrument Formatting

Based on the feedback from the cognitive interviews and pretesting, we made modifications to the wording of 4 items and added explanations to 2 items in order to make them easier to understand. This self-administered measurement instrument with 29 items was used to collected survey data.

Results of Survey

Characteristics of Survey Respondents

Survey responses were collected from a total of 201 respondents (Multimedia Appendix 3) from 3 hospitals in Shanghai, China, of which 156 responses (77.6%) were valid. No data were missing. The ratio of participants to items was 5.4 to 1.

Reduction of Items for the Measurement Instrument

One item—usage behavior—was deleted based on item-scale correlation, corrected item-to-total correlation, and effect on Cronbach-α-if-the-item-was-deleted criteria (Table 1).

Latent Construct of the Measurement Instrument

Exploratory factor analysis was deemed to be appropriate (Kaiser-Meyer-Olkin .923; χ_{378}^2=3859.495; and significant Bartlett test of sphericity, P<.001). Eight components, which explained 80.6% of the variance, were extracted (Table 2; Multimedia Appendix 4; Multimedia Appendix 5). For interpretability, we classified decision change, process change and outcome change as one factor—*Perceived benefit*—thereby, the constructs of measurement instrument reflected the 6 variables in the hypothesis model.

Table 2. Principal component analysis results.

Component	Extraction			Rotation
	Sums of squared loadings	Variance (%)	Cumulative variance (%)	Sums of squared loadings
Perceived ease of use	14.447	51.596	51.596	11.354
System quality	2.504	8.941	60.537	9.824
Information quality	1.423	5.082	65.620	11.299
Service quality	1.212	4.328	69.948	5.687
Decision change	0.841	3.005	72.953	6.449
Process change	0.779	2.780	75.733	7.736
Outcome change	0.715	2.555	78.288	6.588
Acceptance	0.658	2.350	80.638	5.997

Reliability and Validity of Measurement Instrument

The 28-item scale appeared to be internally consistent (Cronbach α=.963). The Cronbach α for the 6 subscales ranged from .760 to .949. Content validity of the overall scale was 0.943. Values of average variance extracted ranged from .582 to .756 and met the >.50 restrictive criterion, which indicated acceptable convergent validity. The values of heterotrait-monotrait ratio ranged from 0.376 to 0.896 and met the <0.90 restrictive criterion, which indicated acceptable discriminant validity of constructs (Table 3, Multimedia Appendix 6).

Table 3. Internal consistency, convergent validity, and discriminant validity of constructs.

Variables	Heterotrait-monotrait ratio						Average variance extracted	Composite reliability
	Perceived ease of use	System quality	Information quality	Service quality	Perceived benefit	Acceptance		
Perceived ease of use	1	0.753	0.765	0.412	0.657	0.736	.582	.892
System quality	0.753	1	0.637	0.376	0.455	0.636	.674	.803
Information quality	0.765	0.637	1	0.721	0.729	0.767	.620	.760
Service quality	0.412	0.376	0.721	1	0.654	0.673	.752	.858
Perceived benefit	0.657	0.455	0.729	0.654	1	0.896	.595	.935
Acceptance	0.736	0.636	0.767	0.673	0.896	1	.756	.949

Model Validation

Hypothesized Model Modification

The chi-square of the hypothesized model was significant (χ_{30}^2=126.962, P<.001; ratio of chi-square over degrees of freedom 4.232). Model fit indices (comparative fit index 0.921; goodness-of-fit index 0.874; root mean square error of approximation 0.144; standardized root mean square residual 0.131) suggested the hypothesized model needed to be modified in order to have a better fitting model: 2 paths, predicting *Acceptance* from *Information quality* and *Service quality*, were added, and one path, predicting *Perceived ease of use* from *Service quality*, was moved, which significantly improved the model and lowered the chi-square values. This meant that in addition to the relationship between *Perceived ease of use* and *Information quality* or *Acceptance*, there was also a direct relationship between *Information quality* and *Acceptance*.

Revised Model Fit and Pathway Coefficients

The chi-square of the revised model was not significant (χ_{26}^2=36.984, P=.08; ratio of chi-square over degrees of freedom 1.422). Model fit indices (comparative fit index 0.991; goodness-of-fit index 0.957; root mean square error of approximation 0.052; standardized root mean square residual 0.028) indicated a good-fitting model (Figure 2). All of the path coefficients between measured variables and factors in the final model were significant (2-tailed, P<.05). Better *System quality* (P<.001) and better *Information quality* (P<.001) significantly increased *Perceived ease of use*. Better *Information quality* (P=.04), better *Service quality* (P<.001), and *Perceived ease of use* (P<.001) significantly increased *Acceptance*. *Acceptance* and *Perceived benefit* were interrelated (Figure 2, Table 4). Variables in the final model accounted for 89% of the variance in *Acceptance* (Table 5). Parameter estimation of error in measurement, standardized total effects, direct effects, and indirect effects are shown in Multimedia Appendix 7-10.

Figure 2. Final evaluation model (comparative fit index 0.991; goodness-of-fit index 0.957; root mean square error of approximation 0.052; standardized root mean square residual 0.028).

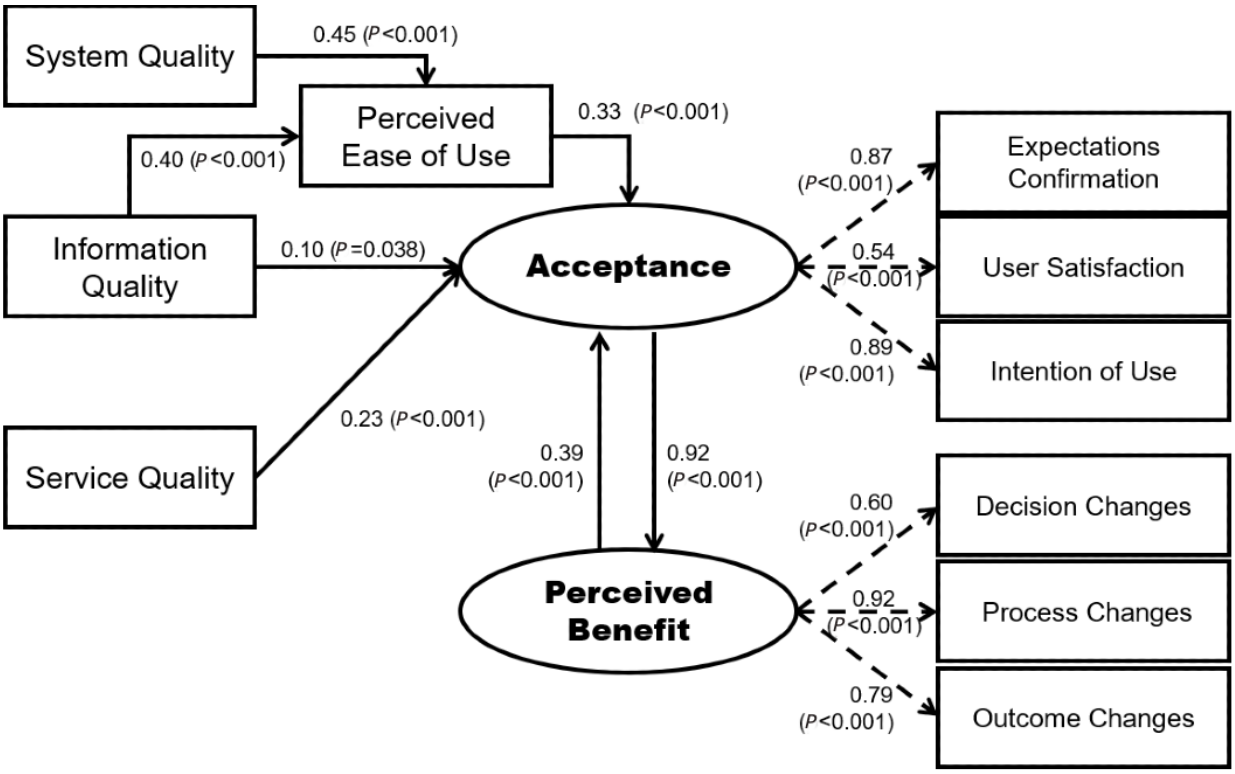


Table 4. Parameter estimation for path coefficients.

Pathway		Regression weights	Standardized regression weights	Standard error	Critical ratio	*P* value
Perceived ease of use	System quality	0.292	0.446	0.041	7.139	<.001
Perceived ease of use	Information quality	0.378	0.405	0.058	6.484	<.001
Acceptance	Information quality	0.117	0.099	0.057	2.070	.04
Acceptance	Service quality	0.235	0.232	0.052	4.525	<.001
Acceptance	Perceived ease of use	0.413	0.325	0.084	4.933	<.001
Expectations confirmation	Acceptance	1	0.866	N/A[a]	N/A	N/A
User satisfaction	Acceptance	0.522	0.536	0.072	7.241	<.001
Intention of use	Acceptance	0.981	0.893	0.062	15.804	<.001
Decision change	Benefit	1	0.595	N/A	N/A	N/A
Process change	Benefit	1.274	0.923	0.161	7.935	<.001
Outcome change	Benefit	1.182	0.788	0.157	7.507	<.001
Benefit	Acceptance	0.599	0.925	0.078	7.657	<.001
Acceptance	Benefit	0.599	0.388	0.078	7.657	<.001

[a] N/A: not applicable.

Table 5. Squared multiple correlations.

Variables	Estimate
Perceived ease of use	0.538
Benefit	0.932
Outcome change	0.621
Process change	0.851
Decision change	0.491
Acceptance	0.89
Expectations confirmation	0.75
Intention of use	0.797
User satisfaction	0.853

Discussion

Main Findings

User acceptance was established as central to AI-enabled clinical decision support system success in the evaluation framework. A 28-item measurement instrument was evaluated, yielding an instrument that quantifies 6 variables: *System quality*, *Information quality*, *Service quality*, *Perceived ease of use*, *User acceptance*, and *Perceived benefit*.

User Acceptance is the Central Dimension

User acceptance is the traditional focus of evaluation in determining the success of an information system [15,17,32]. User acceptance is a synthesized concept—we used expectation confirmation, user satisfaction, and intention of use as secondary indicators. The item *system usage* was removed; DeLone and McLean [38] suggested that “intention to use,” that is, *intention of use* in our study, may be a worthwhile alternative measure in some contexts. Our work demonstrated that the use or nonuse of AI-enabled clinical decision support systems is not a universal success criterion. Therefore, the item was removed from the measurement instrument. The nature of health care settings, wherein diverse perspectives, power asymmetry, and politically led changes co-exist, supports this approach [26]. The use of an AI-enabled clinical decision support system tends to be mandatory, thus it is difficult to interpret users’ evaluations with respect to system usage. Our model demonstrated that User *Acceptance* of AI-enabled clinical decision support systems was directly determined by *Perceived ease of use*, *Information quality*, *Service quality*, and *Perceived benefit*.

Perceived Ease of Use

In this study, perceived ease of use encompassed human–computer interaction (eg, user interface, data entry, information display, legibility, response time), ease of learning, and workflow integration [17,56,57]. *Perceived ease of use* was a mediation variable between *System quality*, *Information quality*, and *Acceptance*. *System quality* did not directly affect user *Acceptance*, but indirectly exerted influence through

Perceived ease of use, principally because clinicians' intuitive feelings of ease of use are fixed on external, tangible, and accessible features. Engineering-oriented performance characteristics of an AI-enabled clinical decision support system and necessary supporting functionalities are not their main concerns.

Information Quality

Information quality refers to reliable and valid suggestions, provided by an AI-enabled clinical decision support systems, and directly and indirectly affected user *Acceptance*. Suggestions without reliability or validity not only reflects low diagnostic performance of AI-enabled clinical decision support systems but also may excessively interrupt daily work [36,58], negatively affecting ease of use and further lowering user acceptance.

Service Quality

Service quality required by clinicians emphasizes knowledge updating for timeliness and system improvement [9,56].

Perceived Benefit

Perceived benefit and user *Acceptance* were interrelated; and clinicians are always concerned with the usefulness of AI-enabled clinical decision support system adoption for themselves, groups, and patients [19]. AI-enabled clinical decision support system products with anticipated benefits are more likely to be accepted by clinicians. As demonstrated in our study, *Perceived benefit* was not the conclusive criterion of AI-enabled clinical decision support system success even if it could be measured with precision [59]. There will be a comparison between assumptions and expectations of personal preference with perceived benefit [36]. When clinicians are not willing to accept a new AI-enabled clinical decision support systems, the system will face adoption difficulties in clinical practice even if the system is considered to be a benefit to quality of care and patients' outcomes in general.

Recommendations of Benefit Measures for AI-Enabled Clinical Decision Support Systems

Decision Changes

We recommend using *Decision change* as an outcome measure rather than appropriate decisions. Decision change for AI-enabled clinical decision support system usage underlines decision inconsistency between system and human. These decision-making suggestions might correct users' clinical orders, particularly for those who have insufficient practical experience [21]. Consequently, measuring user decision change (eg, tests cancel, order optimization) is more straightforward than measuring appropriate decisions.

Process Changes

Process change, which is similar to perceived usefulness [39], mainly covers individual, group, or organization levels of performance improvement. This study used knowledge, skills, confidence [17,25,60-62], and work efficiency [17,61] as indicators of individual performance and used quality of health care and documentation [57,62-66] as indicators of group or organization performance.

Outcome Changes

Outcome measures tended to be complicated indicators of AI-enabled clinical decision support system success, which often failed to be objective in clinical settings [15,58]. Beneficial patient outcomes from AI-enabled clinical decision support system implementations are the concern of all stakeholders. But there remains a paucity of high-quality evidence for outcome measures [19]. Consequently, although both subjective and objective measures of AI-enabled clinical decision support system success should compensate for the shortcomings of each other, our work showed that it is valuable to evaluate clinicians' attitude toward perceived benefit for patients that can be obtained from specific AI-enabled clinical decision support system implementation under the health care contexts when objective measures are difficult to qualify.

Limitations

This study is an innovative attempt and pilot examination of an evaluation framework in relation to AI-enabled clinical decision support system success. This evaluation framework is widely applicable, with a broad scope in clinically common and multidisciplinary interoperable scenarios. In order to test the validity of the variables and the hypotheses about their relationships, an empirical methodology was needed. Specifically, the items of the measurement instrument were developed targeting diagnostic AI-enabled clinical decision support systems, and AI-enabled clinical decision support systems designed to support the risk assessment of the venous thromboembolism among inpatients was the focus. Thus, one potential limitation may arise due to this narrow focus. A future expanded evaluation framework would require validation among diverse populations and encompassing AI-enabled clinical decision support systems with diverse functions.

Implications and Conclusion

This study offers unique insight into AI-enabled clinical decision support system evaluation from a user-centric perspective, and the evaluation framework can support stakeholders to understand user acceptance of AI-enabled clinical decision support system products with various functionalities. Given the commonality and interoperability of this evaluation framework, it is widely applicable in different implementations, that is, this framework can be used to evaluate success of various AI-enabled clinical decision support systems.

From a theoretical point of view, this framework can be an evaluation approach to help in describing and understanding AI-enabled clinical decision support system success with a user acceptance–centric evaluation process. There are also practical implications in terms of how this evaluation framework is applied in clinical settings. The 28-item diagnostic AI-enabled clinical decision support system success measurement instrument, divided into 6 model variables, showed good psychometric qualities. The measurement instrument can be a useful resource for health care organizations or academic institutions designing and conducting evaluation projects on specific AI-enabled clinical decision support systems. At the same time, if the measurement instrument is to be used for AI-enabled clinical decision support system products with different functionalities in a specific scenario, item modifications, cross-cultural adaptation, and tests of reliability and validity testing (in accordance with scale development guidelines [52]) is needed.

Acknowledgments

This work was supported by the Doctoral Innovation Fund in Shanghai Jiao Tong University School of Medicine 2019 [BXJ201906]; the Shanghai Municipal Education Commission-Gaoyuan Nursing Grant Support [Hlgy1906dxk]; and the Shanghai Municipal Commission of Health and Family Planning (Grant No. 2018ZHYL0223).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Evaluation target of model variables.
[DOCX File , 36 KB - jmir_v23i6e25929_app1.docx]

Multimedia Appendix 2
Characteristics of the Delphi expert panel.
[DOCX File , 24 KB - jmir_v23i6e25929_app2.docx]

Multimedia Appendix 3
Sociodemographic characteristics of respondents.
[DOCX File , 23 KB - jmir_v23i6e25929_app3.docx]

Multimedia Appendix 4
Structure matrix of measurement instrument.
[DOCX File , 21 KB - jmir_v23i6e25929_app4.docx]

Multimedia Appendix 5
Component correlation matrix.
[DOCX File , 24 KB - jmir_v23i6e25929_app5.docx]

Multimedia Appendix 6
Standardized factor loading of the measurement instrument.
[DOCX File , 28 KB - jmir_v23i6e25929_app6.docx]

Multimedia Appendix 7
Parameter estimation of error in measurement.
[DOCX File , 24 KB - jmir_v23i6e25929_app7.docx]

Multimedia Appendix 8
Standardized total effects.
[DOCX File , 24 KB - jmir_v23i6e25929_app8.docx]

Multimedia Appendix 9
Standardized direct effects.
[DOCX File , 24 KB - jmir_v23i6e25929_app9.docx]

Multimedia Appendix 10
Standardized indirect effects.
[DOCX File , 24 KB - jmir_v23i6e25929_app10.docx]

References

1. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc 2001;8(6):527-534 [FREE Full text] [Medline: 11687560]

2. Haynes RB, Wilczynski NL, Computerized Clinical Decision Support System (CCDSS) Systematic Review Team. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: methods of a decision-maker-researcher partnership systematic review. Implement Sci 2010 Feb 05;5:12 [FREE Full text] [doi: 10.1186/1748-5908-5-12] [Medline: 20181104]
3. Grout RW, Cheng ER, Carroll AE, Bauer NS, Downs SM. A six-year repeated evaluation of computerized clinical decision support system user acceptability. Int J Med Inform 2018 Apr;112:74-81 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.01.011] [Medline: 29500025]
4. Jia P, Jia P, Chen J, Zhao P, Zhang M. The effects of clinical decision support systems on insulin use: a systematic review. J Eval Clin Pract 2020 Aug;26(4):1292-1301. [doi: 10.1111/jep.13291] [Medline: 31782586]
5. Daniel G, Silcox C, Sharma I, Wright M. Current state and near-term priorities for ai-enabled diagnostic support software in health care. Duke Margolis Center for Health Policy. 2019. URL: https://healthpolicy.duke.edu/sites/default/files/2019-11/dukemargolisaienableddxss.pdf [accessed 2019-11-19]
6. Salem H, Attiya G, El-Fishawy N. A survey of multi-agent based intelligent decision support system for medical classification problems. Int J Comput Appl 2015 Aug 18;123(10):20-25. [doi: 10.5120/ijca2015905529]
7. Aljaaf A, Al-Jumeily D, Hussain A, Fergus P, Al-Jumaily M, Abdel-Aziz K. Toward an optimal use of artificial intelligence techniques within a clinical decision support system. 2015 Presented at: Science and Information Conference; July 28-30; London, UK. [doi: 10.1109/sai.2015.7237196]
8. Richard A, Mayag B, Talbot F, Tsoukias A, Meinard Y. What does it mean to provide decision support to a responsible and competent expert? Euro J Decis Process 2020 Aug 12;8(3-4):205-236. [doi: 10.1007/s40070-020-00116-7]
9. Faviez C, Chen X, Garcelon N, Neuraz A, Knebelmann B, Salomon R, et al. Diagnosis support systems for rare diseases: a scoping review. Orphanet J Rare Dis 2020 Apr 16;15(1):94 [FREE Full text] [doi: 10.1186/s13023-020-01374-z] [Medline: 32299466]
10. Wulff A, Montag S, Marschollek M, Jack T. Clinical decision-support systems for detection of systemic inflammatory response syndrome, sepsis, and septic shock in critically ill patients: a systematic review. Methods Inf Med 2019 Dec;58(S 02):e43-e57 [FREE Full text] [doi: 10.1055/s-0039-1695717] [Medline: 31499571]
11. Langerhuizen DWG, Janssen SJ, Mallee WH, van den Bekerom MPJ, Ring D, Kerkhoffs GMMJ, et al. What are the applications and limitations of artificial intelligence for fracture detection and classification in orthopaedic trauma imaging? a systematic review. Clin Orthop Relat Res 2019 Nov;477(11):2482-2491 [FREE Full text] [doi: 10.1097/CORR.0000000000000848] [Medline: 31283727]
12. Yassin NIR, Omran S, El Houby EMF, Allam H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. Comput Methods Programs Biomed 2018 Mar;156:25-45. [doi: 10.1016/j.cmpb.2017.12.012] [Medline: 29428074]
13. Ferrante di Ruffano L, Takwoingi Y, Dinnes J, Chuchu N, Bayliss SE, Davenport C, Cochrane Skin Cancer Diagnostic Test Accuracy Group. Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults. Cochrane Database Syst Rev 2018 Dec 04;12:CD013186 [FREE Full text] [doi: 10.1002/14651858.CD013186] [Medline: 30521691]
14. Roumeliotis N, Sniderman J, Adams-Webber T, Addo N, Anand V, Rochon P, et al. Effect of electronic prescribing strategies on medication error and harm in hospital: a systematic review and meta-analysis. J Gen Intern Med 2019 Oct;34(10):2210-2223 [FREE Full text] [doi: 10.1007/s11606-019-05236-8] [Medline: 31396810]
15. Rawson TM, Moore LSP, Hernandez B, Charani E, Castro-Sanchez E, Herrero P, et al. A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately? Clin Microbiol Infect 2017 Aug;23(8):524-532 [FREE Full text] [doi: 10.1016/j.cmi.2017.02.028] [Medline: 28268133]
16. Oluoch T, Santas X, Kwaro D, Were M, Biondich P, Bailey C, et al. The effect of electronic medical record-based clinical decision support on HIV care in resource-constrained settings: a systematic review. Int J Med Inform 2012 Oct;81(10):e83-e92. [doi: 10.1016/j.ijmedinf.2012.07.010] [Medline: 22921485]
17. Carter J, Sandall J, Shennan AH, Tribe RM. Mobile phone apps for clinical decision support in pregnancy: a scoping review. BMC Med Inform Decis Mak 2019 Nov 12;19(1):219 [FREE Full text] [doi: 10.1186/s12911-019-0954-1] [Medline: 31718627]
18. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SAMD) - discussion paper and request for feedback. US Food and Drug Administration. 2019. URL: https://www.fda.gov/media/122535/download [accessed 2020-10-05]
19. Herasevich V, Pickering B. Health Information Technology Evaluation Handbook: From Meaningful Use to Meaningful Outcome. Boca Raton, Florida: CRC Press, Taylor & Francis Group; 2018.
20. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating ai and practical implications. Yearb Med Inform 2019 Aug;28(1):128-134 [FREE Full text] [doi: 10.1055/s-0039-1677903] [Medline: 31022752]
21. Cresswell K, Callaghan M, Mozaffar H, Sheikh A. NHS Scotland's decision support platform: a formative qualitative evaluation. BMJ Health Care Inform 2019 May;26(1):e100022 [FREE Full text] [doi: 10.1136/bmjhci-2019-100022] [Medline: 31160318]

22. Reis WC, Bonetti AF, Bottacin WE, Reis AS, Souza TT, Pontarolo R, et al. Impact on process results of clinical decision support systems (CDSSs) applied to medication use: overview of systematic reviews. Pharm Pract (Granada) 2017;15(4):1036 [FREE Full text] [doi: 10.18549/PharmPract.2017.04.1036] [Medline: 29317919]
23. Scott PJ, Brown AW, Adedeji T, Wyatt JC, Georgiou A, Eisenstein EL, et al. A review of measurement practice in studies of clinical decision support systems 1998-2017. J Am Med Inform Assoc 2019 Oct 01;26(10):1120-1128 [FREE Full text] [doi: 10.1093/jamia/ocz035] [Medline: 30990522]
24. Blum D, Raj SX, Oberholzer R, Riphagen II, Strasser F, Kaasa S, EURO IMPACT, European Intersectorial Multidisciplinary Palliative Care Research Training. Computer-based clinical decision support systems and patient-reported outcomes: a systematic review. Patient 2015 Oct 29;8(5):397-409. [doi: 10.1007/s40271-014-0100-1] [Medline: 25432150]
25. Souza NM, Sebaldt RJ, Mackay JA, Prorok JC, Weise-Kelly L, Navarro T, CCDSS Systematic Review Team. Computerized clinical decision support systems for primary preventive care: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. Implement Sci 2011 Aug 03;6:87 [FREE Full text] [doi: 10.1186/1748-5908-6-87] [Medline: 21824381]
26. Mussi CC, do Valle Pereira CD, de Oliveira Lacerda RT, dos Santos EM. Pre-implementation evaluation of a nationwide information system for university hospitals: lessons learned from a study in Brazil. Behav Inf Technol 2018 Feb 05;37(3):217-231. [doi: 10.1080/0144929x.2018.1430850]
27. Lau F, Kuziemsky C. Handbook of eHealth Evaluation: An Evidence-based Approach. Victoria, Canada: University of Victoria; 2016.
28. Cresswell K, Williams R, Sheikh A. Developing and applying a formative evaluation framework for health information technology implementations: qualitative investigation. J Med Internet Res 2020 Jun 10;22(6):e15068 [FREE Full text] [doi: 10.2196/15068] [Medline: 32519968]
29. Scott P, Keizer N, Georgiou A. Applied Interdisciplinary Theory in Health Informatics: A Knowledge Base for Practitioners. USA: IOS Press; 2019.
30. Abraham C, Boudreau M, Junglas I, Watson R. Enriching our theoretical repertoire: the role of evolutionary psychology in technology acceptance. Eur J Inf Syst 2017 Dec 19;22(1):56-75. [doi: 10.1057/ejis.2011.25]
31. Trinkley KE, Kahn MG, Bennett TD, Glasgow RE, Haugen H, Kao DP, et al. Integrating the practical robust implementation and sustainability model with best practices in clinical decision support design: implementation science approach. J Med Internet Res 2020 Oct 29;22(10):e19676 [FREE Full text] [doi: 10.2196/19676] [Medline: 33118943]
32. Al-Gahtani SS, King M. Attitudes, satisfaction and usage: Factors contributing to each in the acceptance of information technology. Behav Inf Technol 1999 Jan;18(4):277-297. [doi: 10.1080/014492999119020]
33. Jones S, Hughes J. Understanding IS evaluation as a complex social process: a case study of a UK local authority. Eur J Inf Syst 2017 Dec 19;10(4):189-203. [doi: 10.1057/palgrave.ejis.3000405]
34. Walsh I. Using quantitative data in mixed-design grounded theory studies: an enhanced path to formal grounded theory in information systems. Eur J Inf Syst 2017 Dec 19;24(5):531-557. [doi: 10.1057/ejis.2014.23]
35. Sligo J, Gauld R, Roberts V, Villa L. A literature review for large-scale health information system project planning, implementation and evaluation. Int J Med Inform 2017 Dec;97:86-97. [doi: 10.1016/j.ijmedinf.2016.09.007] [Medline: 27919399]
36. Cajander Å, Grünloh C. Electronic health records are more than a work tool: conflicting needs of direct and indirect stakeholders. 2019 Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems - CHI '19; May 4-5; Glasgow, Scotland, UK p. 1-13. [doi: 10.1145/3290605.3300865]
37. Ji M, Yu G, Xi H, Xu T, Qin Y. Measures of success of computerized clinical decision support systems: an overview of systematic reviews. Health Policy Technol 2021 Mar;10(1):196-208. [doi: 10.1016/j.hlpt.2020.11.001]
38. Delone WH, McLean ER. The DeLone and McLean model of information systems success: a ten-year update. J Manag Inf Syst 2014 Dec 23;19(4):9-30. [doi: 10.1080/07421222.2003.11045748]
39. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Quarterly 2001 Sep;25(3):351. [doi: 10.2307/3250921]
40. Bhattacherjee A. An empirical analysis of the antecedents of electronic commerce service continuance. Decis Support Syst 2001 Dec;32(2):201-214. [doi: 10.1016/s0167-9236(01)00111-7]
41. Bossen C, Jensen LG, Udsen FW. Evaluation of a comprehensive EHR based on the DeLone and McLean model for IS success: approach, results, and success factors. Int J Med Inform 2013 Oct;82(10):940-953. [doi: 10.1016/j.ijmedinf.2013.05.010] [Medline: 23827768]
42. Tilahun B, Fritz F. Comprehensive evaluation of electronic medical record system use and user satisfaction at five low-resource setting hospitals in ethiopia. JMIR Med Inform 2015 May 25;3(2):e22 [FREE Full text] [doi: 10.2196/medinform.4106] [Medline: 26007237]
43. Tubaishat A. Evaluation of electronic health record implementation in hospitals. Comput Inform Nurs 2017 Jul;35(7):364-372. [doi: 10.1097/CIN.0000000000000328] [Medline: 28221186]
44. Saghaeiannejad-Isfahani S, Saeedbakhsh S, Jahanbakhsh M, Habibi M. Analysis of the quality of hospital information systems in Isfahan teaching hospitals based on the DeLone and McLean model. J Educ Health Promot 2015;4:5 [FREE Full text] [doi: 10.4103/2277-9531.151883] [Medline: 25767816]

45. Abimbola S, Keelan S, Everett M, Casburn K, Mitchell M, Burchfield K, et al. The medium, the message and the measure: a theory-driven review on the value of telehealth as a patient-facing digital health innovation. Health Econ Rev 2019 Jul 03;9(1):21 [FREE Full text] [doi: 10.1186/s13561-019-0239-5] [Medline: 31270685]
46. Hung H, Altschuld JW, Lee Y. Methodological and conceptual issues confronting a cross-country Delphi study of educational program evaluation. Eval Program Plann 2008 May;31(2):191-198. [doi: 10.1016/j.evalprogplan.2008.02.005] [Medline: 18403018]
47. Reio TG, Shuck B. Exploratory factor analysis. Adv Dev Hum Resour 2014 Nov 28;17(1):12-25. [doi: 10.1177/1523422314559804]
48. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychol Bull 1988 May;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
49. Pallant J. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS 4th Edition. Australia: Allen & Unwin; 2011.
50. Burns KEA, Duffett M, Kho ME, Meade MO, Adhikari NKJ, Sinuff T, et al. A guide for the design and conduct of self-administered surveys of clinicians. CMAJ 2008 Jul 29;179(3):245-252 [FREE Full text] [doi: 10.1503/cmaj.080372] [Medline: 18663204]
51. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res 1991 Sep;2(3):192-222. [doi: 10.1287/isre.2.3.192]
52. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149 [FREE Full text] [doi: 10.3389/fpubh.2018.00149] [Medline: 29942800]
53. Spicer J. Making Sense of Multivariate Data Analysis: An Intuitive Approach. London: Sage; 2005.
54. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
55. Tabachnick B, Fidell L. Using Multivariate Statistics. Boston: Pearson Education Inc; 2007:0205459382.
56. Kilsdonk E, Peute LW, Jaspers MWM. Factors influencing implementation success of guideline-based clinical decision support systems: a systematic review and gaps analysis. Int J Med Inform 2017 Dec;98:56-64. [doi: 10.1016/j.ijmedinf.2016.12.001] [Medline: 28034413]
57. Stultz JS, Nahata MC. Computerized clinical decision support for medication prescribing and utilization in pediatrics. J Am Med Inform Assoc 2012;19(6):942-953 [FREE Full text] [doi: 10.1136/amiajnl-2011-000798] [Medline: 22813761]
58. Arts DL, Medlock SK, van Weert HCPM, Wyatt JC, Abu-Hanna A. Acceptance and barriers pertaining to a general practice decision support system for multiple clinical conditions: A mixed methods evaluation. PLoS One 2018;13(4):e0193187 [FREE Full text] [doi: 10.1371/journal.pone.0193187] [Medline: 29672521]
59. Seddon PB. A respecification and extension of the DeLone and McLean model of IS success. Inf Syst Res 1997 Sep;8(3):240-253. [doi: 10.1287/isre.8.3.240]
60. Fathima M, Peiris D, Naik-Panvelkar P, Saini B, Armour CL. Effectiveness of computerized clinical decision support systems for asthma and chronic obstructive pulmonary disease in primary care: a systematic review. BMC Pulm Med 2014 Dec 02;14:189 [FREE Full text] [doi: 10.1186/1471-2466-14-189] [Medline: 25439006]
61. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 3;157(1):29-43. [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
62. Roshanov PS, Misra S, Gerstein HC, Garg AX, Sebaldt RJ, Mackay JA, et al. Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review. Implement Sci 2011 Aug 03;6:92 [FREE Full text] [doi: 10.1186/1748-5908-6-92] [Medline: 21824386]
63. Heselmans A, Van de Velde S, Donceel P, Aertgeerts B, Ramaekers D. Effectiveness of electronic guideline-based implementation systems in ambulatory care settings - a systematic review. Implementation Sci 2009 Dec 30;4(1):1-12. [doi: 10.1186/1748-5908-4-82]
64. Robertson J, Walkom E, Pearson S, Hains I, Williamsone M, Newby D. The impact of pharmacy computerised clinical decision support on prescribing, clinical and patient outcomes: a systematic review of the literature. Int J Pharm Pract 2010 Apr;18(2):69-87. [Medline: 20441116]
65. Shojania KG, Jennings A, Mayhew A, Ramsay CR, Eccles MP, Grimshaw J. The effects of on-screen, point of care computer reminders on processes and outcomes of care. Cochrane Database Syst Rev 2009 Jul 08(3):CD001096 [FREE Full text] [doi: 10.1002/14651858.CD001096.pub2] [Medline: 19588323]
66. Sahota N, Lloyd R, Ramakrishna A, Mackay JA, Prorok JC, Weise-Kelly L, et al. Computerized clinical decision support systems for acute care management: a decision-maker-researcher partnership systematic review of effects on process of care and patient outcomes. Implement Sci 2011;6:91 [FREE Full text] [doi: 10.1186/1748-5908-6-91] [Medline: 21824385]

Abbreviations

AI: artificial intelligence

Edited by R Kukafka; submitted 20.11.20; peer-reviewed by J Li, M Bowden; comments to author 30.12.20; revised version received 12.01.21; accepted 30.04.21; published 02.06.21.

Please cite as:
Ji M, Genchev GZ, Huang H, Xu T, Lu H, Yu G
Evaluation Framework for Successful Artificial Intelligence–Enabled Clinical Decision Support Systems: Mixed Methods Study
J Med Internet Res 2021;23(6):e25929
URL: https://www.jmir.org/2021/6/e25929
doi:10.2196/25929
PMID:34076581

Original Paper

Using the Computer-based Health Evaluation System (CHES) to Support Self-management of Symptoms and Functional Health: Evaluation of Hematological Patient Use of a Web-Based Patient Portal

Jens Lehmann[1,2], MSc; Petra Buhl[2], MSc; Johannes M Giesinger[1], PhD; Lisa M Wintner[3], MA; Monika Sztankay[1], PhD; Lucia Neppl[1], MSc; Wolfgang Willenbacher[2,4], PhD; Roman Weger[2], MA; Walpurga Weyrer[5], MD; Gerhard Rumpold[6,7], PhD; Bernhard Holzner[1,7], PhD

[1]University Hospital of Psychiatry II, Medical University of Innsbruck, Innsbruck, Austria
[2]Oncotyrol – Center for Personalized Cancer Medicine, Innsbruck, Austria
[3]University Hospital of Psychiatry I, Medical University of Innsbruck, Innsbruck, Austria
[4]Internal Medicine V: Haematology & Oncology, Innsbruck University Hospital, Innsbruck, Austria
[5]Comprehensive Cancer Center Innsbruck, Innsbruck, Austria
[6]University Clinic of Medical Psychology, Medical University of Innsbruck, Innsbruck, Austria
[7]Evaluation Software Development GmbH, Innsbruck, Austria

Corresponding Author:
Jens Lehmann, MSc
University Hospital of Psychiatry II
Medical University of Innsbruck
Anichstraße 35
Innsbruck, AT-6020
Austria
Phone: 43 5050481551
Email: jens.lehmann@i-med.ac.at

Abstract

Background: Patient portals offer the possibility to assess patient-reported outcome measures (PROMs) remotely, and first evidence has demonstrated their potential benefits.

Objective: In this study, we evaluated patient use of a web-based patient portal that provides patient information and allows online completion of PROMs. A particular focus was on patient motivation for (not) using the portal. The portal was developed to supplement routine monitoring at the Department of Internal Medicine V in Innsbruck.

Methods: We included patients with multiple myeloma and chronic lymphocytic leukemia who were already participating in routine monitoring at the hospital for use of the patient portal. Patients were introduced to the portal and asked to complete questionnaires prior to their next hospital visits. We used system access logs and 3 consecutive semistructured interviews to analyze patient use and evaluation of the portal.

Results: Between July 2017 and August 2020, we approached 122 patients for participation in the study, of whom 83.6% (102/122) consented to use the patient portal. Patients were on average 60 (SD 10.4) years old. Of patients providing data at all study time points, 37% (26/71) consistently used the portal prior to their hospital visits. The main reason for not completing PROMs was forgetting to do so in between visits (25/84, 29%). During an average session, patients viewed 5.3 different pages and spent 9.4 minutes logged on to the portal. Feedback from interviews was largely positive with no patients reporting difficulties navigating the survey and 50% of patients valuing the self-management tools provided in the portal. Regarding the portal content, patients were interested in reviewing their own results and reported high satisfaction with the dynamic self-management advice, also reflected in the high number of clicks on those pages.

Conclusions: Patient portals can contribute to patient empowerment by offering sought-after information and self-management advice. In our study, the majority of our patients were open to using the portal. The low number of technical complaints and average time spent in the portal demonstrate the feasibility of our patient portal. While initial interest was high, long-term use was considerably lower and identified as the main area for improvement. In a next step, we will improve several aspects of the patient portal (eg, including a reminder to visit the portal before the next appointment and closer PROM symptom monitoring via an onconurse).

(J Med Internet Res 2021;23(6):e26022) doi:10.2196/26022

KEYWORDS

quality of life; monitoring; patient portals; multiple myeloma; chronic lymphocytic leukemia; patient-reported outcome measures; eHealth; mHealth

Introduction

Patient-reported outcomes (PRO) are defined as all reports about the health status given directly by the patient without interpretation of the patient's response by a clinician or anyone else [1]. While they have long been used in clinical trials, they have, in recent decades, also progressed to enriching routine clinical care [2,3]. Driven by technological progress and an increased availability and use of the internet in the population [4], it has become easier to incorporate the patient's perspective into clinical care using electronic patient-reported outcome measures (ePROMs).

In oncological care, PROs can support patient-clinician communication [5,6] and aid early detection of symptoms [7,8] and have been linked to a decrease in hospitalization and emergency department visits [9]. Building on the evidence base showing the benefits of PRO use at the hospital, the use of web-based solutions to assess ePROMs outside the hospital has gained traction. In the last decade, the number of web-based patient portals that enable the completion of ePROMs has risen and recent research has demonstrated the potential benefits of patient portals in large-scale clinical trials [10-13]. For example, in a randomized controlled trial by Denis et al [10], web-based symptom monitoring was associated with increased survival compared to standard imaging surveillance following treatment for lung cancer. The authors argue that web-based symptom monitoring may allow for earlier symptom detection and appropriate reaction by health care professionals (HCPs). Moreover, PRO web monitoring can be highly cost-effective [11] and reduce the administrative burden of assessments inside the hospital, as patients can complete ePROMs from home. Finally, remote assessments are especially helpful in an outpatient setting, as assessments conducted on the day of chemotherapy administration at the hospital have shown to systematically underestimate patients' symptom burden associated with treatment [14].

Despite the benefits shown in study settings, and even though detailed guidance on how to implement PROs into clinical practice exists [15,16], electronic PRO monitoring and especially patient portals are still only occasionally adopted in routine clinical practice. This can be attributed to the limited integration into electronic health records, a lack of financial reimbursement for ePROM assessments, and a lack of standardized assessment methods, which hinder implementation in routine care [2]. Patient portals also vary considerably regarding the focus of the implementation and their goals and use of PROs [17]. More research is needed that evaluates the usability and acceptability of different applications in routine practice to extend and strengthen the evidence base in this heterogeneous and evolving field of research.

At the Department of Internal Medicine V in Innsbruck, a patient portal for outpatients with multiple myeloma (MM) and chronic lymphocytic leukemia (CLL) was developed to conduct remote PRO assessments. The primary aims for development of the portal were to enhance patient empowerment, encourage patient engagement with PROs, and reduce the administrative burden of PRO assessments inside the hospital.

In this study, we evaluated patient use of the various components of the portal and aimed to identify patient lack of motivation for not using the patient portal and potential barriers to accessibility.

Methods

Study Design

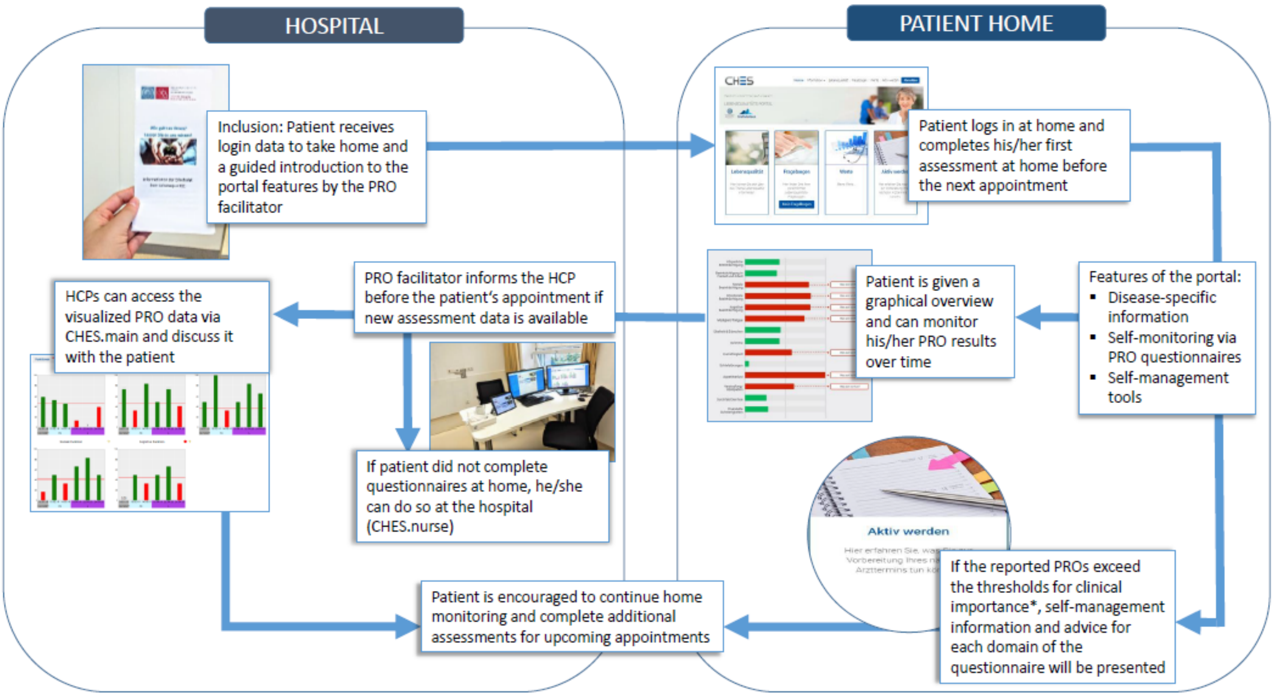
In our observational, longitudinal study, we evaluated patient use of the portal based on two data sources: semistructured interviews conducted during 3 consecutive visits to our unit after introducing eligible patients to the patient portal and system access logs recording the duration of user sessions and how often each page of the web portal was accessed.

Description of the PRO Monitoring and the Patient Portal

In June 2016, we implemented routine ePROM assessments to supplement care of outpatients with MM at the Department of Internal Medicine V in Innsbruck. The two primary aims of the implementation of ePROM assessments were to supplement the Austrian Myeloma Registry (AMR) with PRO data and enrich clinical care with the data [18]. In July 2017, patients with CLL were added as a second patient group. The use of the Computer-based Health Evaluation System (CHES) [19] enables immediate processing and graphical representation of the results. For monitoring at the hospital, patients complete PRO measures before their medical appointment, and the results are presented to the HCP prior to the consultation. The implementation and feasibility of the assessments at the hospital and use of data from patients have been evaluated in the past for the AMR [18], and more details on HCP use of the system are presented elsewhere [20]. This study builds upon our previous implementation strategy and is focused on process evaluation, refinement, and extension (as described in phase IV of the implementation process described by Sztankay et al [18]).

There are 3 main components of CHES: (1) the HCP interface (CHES.main), which presents PRO and patient data to HCPs; (2) the survey interface (CHES.nurse), where the patients complete questionnaires at the hospital; and (3) the patient portal (CHES.portal), which allows remote questionnaire completion and access to supplemental information and self-management advice. See Figure 1 for a visualization of the patient portal and its functionality. See Multimedia Appendix 1 for screenshots of the software and portal.

Figure 1. Computer-Based Health Evaluation System patient portal functions when patients are at home and in the hospital. *Thresholds for clinical importance [21] are used to highlight domains that require discussion with the health care provider. PRO: patient-reported outcome; HCP: healthcare professional.



Inclusion: Patient receives login data to take home and a guided introduction to the portal features by the PRO facilitator

Patient logs in at home and completes his/her first assessment at home before the next appointment

HCPs can access the visualized PRO data via CHES.main and discuss it with the patient

PRO facilitator informs the HCP before the patient's appointment if new assessment data is available

Patient is given a graphical overview and can monitor his/her PRO results over time

Features of the portal:
- Disease-specific information
- Self-monitoring via PRO questionnaires
- Self-management tools

If patient did not complete questionnaires at home, he/she can do so at the hospital (CHES.nurse)

If the reported PROs exceed the thresholds for clinical importance*, self-management information and advice for each domain of the questionnaire will be presented

Patient is encouraged to continue home monitoring and complete additional assessments for upcoming appointments

In this paper, we assess patient use of the patient portal, which features the following functionalities:

- Disease-specific information on CLL and MM: diagnosis, possible symptoms, possible treatments, and links to further information and self-help groups
- PRO assessments with the cancer-specific European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30) in combination with the disease-specific modules EORTC QLQ-MY20 for patients with MM and EORTC QLQ-CLL17 for patients with CLL. The EORTC QLQ-C30 is the most frequently used cancer-specific questionnaire [22] and can be used to measure patient symptoms, functional health, and global quality of life. The modules QLQ-MY20 and QLQ-CLL17 supplement the QLQ-C30 and cover disease-specific issues of quality of life for MM (eg, future perspective, treatment side effects) and CLL patients (eg, symptom burden due to disease and/or treatment, worries/fears regarding health and functioning).
- PRO score review by patient: results are displayed as colored bar charts (longitudinal and cross-sectional). Results that exceed the thresholds for clinical importance [21] are colored red, and results that do not exceed these threshold are colored green (see Lehmann et al [20] for more information).
- Self-management tools and tailored information: based on EORTC QLQ-C30 data, patients are presented with self-management tools for the symptoms and functional health domains. If a patient reports a potentially clinically important result [21], they are directed to the self-management tools (see Lehmann et al [20] for more information). Patients are reminded that in case of severe impairments, they should contact the clinical team directly.

At the outpatient clinic, patients are introduced to the portal by a PRO facilitator (ie, person responsible for the assessments at the hospital). The role of the PRO facilitator is that of a study assistant with a background in psychology or nursing who is trained in the use of PRO data and motivates both patients and HCPs to use the PRO data. Patients are given instructions for use and their personal log-in data for accessing the portal at home by the PRO facilitator.

Typically, patients are advised to complete ePROM assessments in the week before their hospital visits. Visits are scheduled at regular intervals that range between 1 week and 12 months, depending on the disease stage and treatment plan. Patients are encouraged to complete questionnaires as often as they like, even on a daily basis. Questionnaires completed within 7 days of a hospital visit are used to inform the HCP of the patients' health status and linked to the clinical data in the AMR. Data from assessments in between hospital visits are used for research purposes and enable more continuous tracking of patients' health status in the AMR.

Although patients participating in the routine monitoring are encouraged to use the patient portal to report PROs, they can still complete assessments at the hospital (eg, if they forgot to use the portal before their hospital visit).

Study Sample

We defined the following inclusion criteria for participation in monitoring via the patient portal: fluency in German, age 18 years or older, diagnosis of MM or CLL, current treatment at the outpatient clinic of the Department of Internal Medicine V or the Comprehensive Cancer Center Innsbruck of the Medical University of Innsbruck, consent given to routine PRO monitoring at the hospital, and completion of at least one ePROM during a prior visit to the hospital.

Patients were deemed ineligible if they had no access to a computer or the internet or lacked sufficient knowledge to log into a website using a username and password. All patients provided informed consent to use the patient portal. Patients who declined to use the portal were asked for the reason for their refusal. The study and use of patient data are covered by the ethics approval for the AMR issued by the ethics committee of the Medical University of Innsbruck (study number AN3252 266/4.2 386/5.14).

At baseline, demographics (sex, age, marital status, education level, and employment status) and experience with the internet and computer technology (frequency, duration) were collected via a questionnaire. Diagnosis and cancer stage were obtained from the hospital's medical records.

Selection of Outcome Measures

We used different sources to evaluate patient use and perception of the patient portal. Table 1 displays the selection of outcome measures.

Table 1. Selection of outcome measures.

Questions addressed	Outcome measure	Assessment method	Data type
How often do patients use the portal? Is the portal feasible for use during routine clinical care?	ePROM[a] completion rate (number of completed ePROMs before study time points)	Assessed via CHES[b]	Quantitative data
Why do patients use (or not use) the portal? What is their feedback on portal components?	Patient perspectives on portal components and motivation to use (or not use) the portal; accessibility barriers identified	Semistructured interviews	Qualitative and quantitative data/questions
How often do patients log into the portal? How long do patients log into the portal in a single session? Which pages are viewed and how often?	Patient user patterns in the portal: number and duration of log-ins per patient; portal page views	Assessed via CHES portal log data	Quantitative data

[a] ePROM: electronic patient-reported outcome measure.

[b] CHES: Computer-based Health Evaluation System.

Patient Interviews

Each patient participated in 3 semistructured face-to-face interviews conducted by authors LN, PB, and JL following a fixed schedule during 3 consecutive visits to the outpatient unit. See Figure 2 for a flowchart of the interview process. The topics covered in the interviews were as follows:

T1: Guided introduction to the portal and user log-in

- Possible difficulties logging in
- Navigating the portal
- Reviewing PRO results
- Other suggestions/remarks by patients

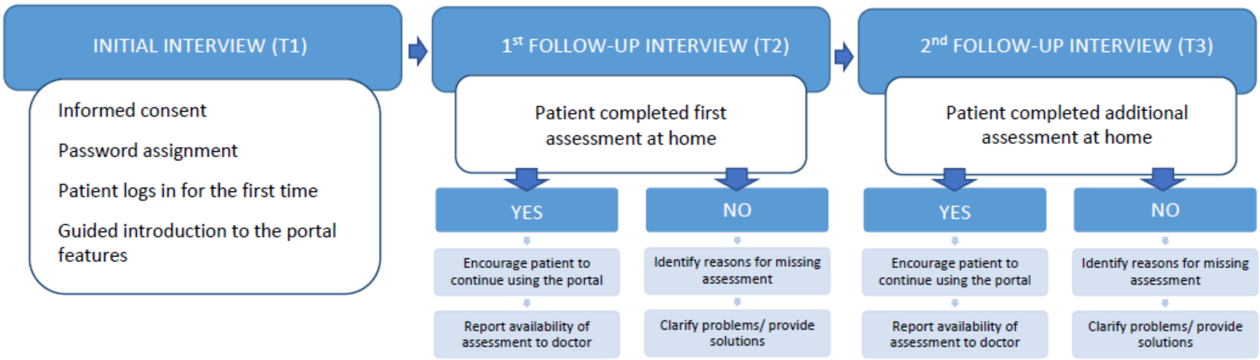
T2: Evaluation of use: acceptability and usability

- Satisfaction with instructions for completion of questionnaire
- Navigating the portal
- Technical difficulties
- Reviewing PRO results and accessing self-help tools
- Relevance and usefulness of provided content

T3: Evaluation of use: acceptability and usability

- Technical difficulties
- Interest in content not explored during previous log-ins
- Satisfaction with design
- Other suggestions/remarks by patients

Figure 2. Patient interview procedure and topics.



ePROM Completion Rate

To evaluate initial interest in the portal, we assessed the proportion of patients in routine monitoring at the hospital consenting to use the patient portal at T1. To evaluate continued use, we assessed the proportion of patients who completed ePROMs prior to hospital visits at T2 and T3; ePROMs were considered to be linked to a hospital visit if they were completed in the 7 days before the visit.

CHES Log Data

The frequency (absolute number) and duration (in minutes) of the use of the website were determined by the CHES system log. These data were gathered irrespective of the interview time points each time the patient logged into the system. Log data were collected per session but were not linked to individual patients. Therefore, patients with more sessions have a greater weight in this analysis. Log-ins that occurred on the same day were considered a single session, and durations of these sessions were summed.

Data Analysis

Sociodemographic and clinical data were analyzed at time point T1. Comparison of patients completing ePROMs via the patient portal and of those who declined to use the portal were made with *t* tests (for parametric variables) and chi-square tests (for nonparametric variables).

The interviewer took field notes during the interviews and for each open question. Responses from the interviews were either analyzed descriptively (for yes/no questions) or paraphrased and category-coded (for open answers) independently by two researchers (JL and PB) and harmonized by discussion in case of different coding. We translated selected quotes into English for the results section of this paper.

We also analyzed the frequency (number) of views for each portal page, excluding views on the start page, where patients are directed automatically after logging in. Further, we calculated the time patients needed to complete the questionnaires online (mean completion time per patient averaged across all patients).

Results

Participant Enrollment and Baseline Characteristics

Recruitment began in July 2017 and was open until August 2020. During the study period, we identified 142 eligible patients already participating in the electronic PRO monitoring in the hospital, of whom 85.9% (122/142) were approached for study participation and use of the patient portal. Of those, 83.6% (102/122) consented to be included in the patient portal. Of 20 patients not willing to use the patient portal, 18 patients stated a preference of questionnaire completion only at the hospital as the reason. The complete enrollment process is shown in Figure 3.

The full sociodemographic information and clinical data are given in Table 2. There were no statistically significant differences regarding age, sex, education, time since the initial diagnosis, and type of internet use between those who agreed and those who refused to participate (all $P>.09$). The age range of patients included in the portal was 39 to 83 years, and the age range of patients not included in the portal was 39 to 77 years. A statistically significant difference was found for general internet use (see Table 2), with participants who consented to use the portal reporting higher internet use than those who did not consent to use the portal: of patients who used the portal, 94% (94/100) reported using the internet at least multiple times per week compared to 76% (17/22) for patients who did not use the portal.

Figure 3. Recruitment flowchart (study patient denotes patients who were also participating in other clinical studies and were not included in our study so as not to overburden the patient with clinical questionnaires).

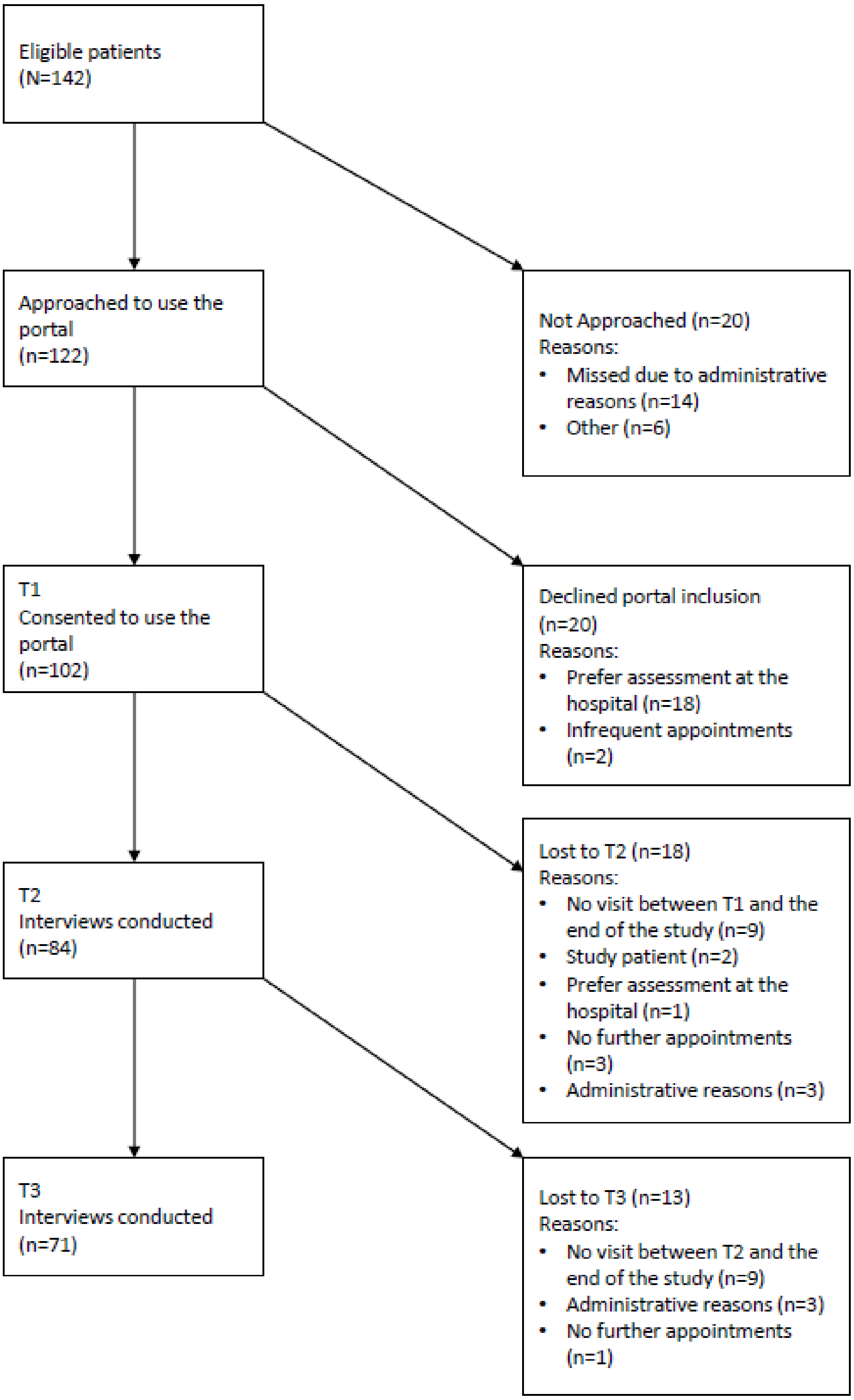


Table 2. Sociodemographic information.

Characteristic	Included in patient portal (n=102)	Not included in patient portal (n=23)	Statistic		*P* value
			χ^2	*t* score	
Sex, n (%)	—[a]	—	0.1	—	.80
Female	37 (37)	9 (39)	—	—	—
Male	65 (64)	14 (61)	—	—	—
Age (years), mean (SD)	59.9 (10.5)	63.2 (10.5)	—	1.33	.19
Diagnosis, n (%)	—	—	4.0	—	.047
Multiple myeloma	63 (62)	9 (39)	—	—	—
Chronic lymphocytic leukemia	39 (38)	14 (61)	—	—	—
Time since diagnosis (years), mean (IQR)	4.5 (0.9-6.8)	4.7 (0.9-6.6)	—	1.73	.09
Highest education, n (%)	—	—	0.9	—	.84
Compulsory or lower	9 (9)	1 (5)	—	—	—
Vocational training	55 (55)	10 (50)	—	—	—
High school certificate	18 (18)	4 (20)	—	—	—
University	18 (18)	5 (25)	—	—	—
Missing data[b]	2	3	—	—	—
Internet use (type), n (%)	—	—	0.7	—	.40
Private use only	66 (65)	17 (74)	—	—	—
Job and private use	36 (35)	6 (26)	—	—	—
Internet use (frequency), n (%)	—	—	10.0	—	.02
>Once per month	2 (2)	0 (0)	—	—	—
Multiple times a month	4 (4)	5 (23)	—	—	—
Multiple times a week	30 (30)	4 (17)	—	—	—
Daily	64 (64)	13 (59)	—	—	—
Missing data[b]	2	1	—	—	—

[a] Not applicable.

[b] Missing values were not included in the calculation of percentages.

Interviews

Interviews lasted between 10 and 30 minutes. Only 5% (5/102) patients required help from the PRO facilitator because they were not able to read their username and password (small font size). After the log-in and during the first interview, no patients reported technical difficulties or had difficulties navigating the portal or the questionnaire. Table 3 shows the number of completed questionnaires before the interviews and the reasons for noncompletion. Of the patients who participated in all 3 interviews, 37% (26/71) completed the questionnaires prior to the T2 and T3 interview, while the others completed the questionnaires only at the hospital visit.

Table 4 shows the use and evaluation of the portal as reported by patients who completed questionnaires prior to the interviews. Patients who completed questionnaires prior to their follow-up appointments reported using the portal in different ways: the percentage of patients reading additional portal content (eg, disease-specific information) declined over time from 71% (27/38) at T2 to 41% (13/32) at T3. The self-management tools were used and deemed useful by 50% (17/34) of patients at the second and 42% (13/31) at the third interview.

Table 3. Questionnaire completion rate and reasons for not completing questionnaires in the patient portal before hospital visits.

Completion rate or reason for noncompletion	Interview 1 (n=102), n (%)	Interview 2 (n=84), n (%)	Interview 3 (n=71), n (%)
Questionnaire completed before appointment			
Yes	102 (100)	38 (45)	32 (45)
No	—[a]	46 (55)	39 (55)
Reasons not completed			
Forgot	—	25 (29)	18 (25)
Technical difficulties	—	9 (11)	0 (0)
Loss of log-in data	—	0 (0)	3 (4)
Study patient[b]	—	0 (0)	1 (1)
Lack of time/motivation	—	5 (6)	6 (9)
Preferred assessment at hospital	—	3 (4)	7 (10)
Other	—	4 (5)	4 (6)

[a] Not applicable.

[b] Study patients were also participating in other clinical studies using the same questionnaires or similar; they sometimes confused questionnaires from other studies with our study's questionnaires.

Table 4. Use and evaluation of the patient portal as reported in the interviews.

Patient-reported behavior and evaluation of the portal	Completed questionnaires, n (%)		
	Interview 1 (n=102)	Interview 2 (n=38)	Interview 3 (n=32)
Reading (additional) portal content			
No	—[a]	27 (71)	13 (41)
Yes	—	11 (29)	18 (56)
Looking at one's own results			
Yes	96 (96)	25 (69)	—
No	4 (4)	11 (31)	—
Missing data[b]	2	2	—
Self-management tools			
Inspected self-management tools and found them to be useful	—	17 (50)	13 (42)
Inspected self-management tools and did not find them useful	—	4 (12)	2 (6)
Did not inspect self-management tools	—	13 (38)	16 (52)
Missing data[b]	—	4	1
Reason for not inspecting self-management tools			
No impairments reported[c]	—	6 (18)	8 (26)
Lack of time	—	1 (3)	1 (3)
Other	—	6 (18)	7 (23)

[a] Not applicable.

[b] Missing values were not included in the calculation of percentages.

[c] If a patient did not report impairments above the thresholds for clinical importance, the software did not suggest viewing self-management tools when looking at their own results.

Patient Comments

Patients were encouraged to provide additional comments following their answers to the interview questions. In the first interview, patients reported high satisfaction with the presentation of the results as bar charts.

> *...bar charts are a good way of presenting the results. I like that I can compare my results to those of other patients with cancer.*
>
> *I like that I can see my results after having completed the questionnaire.*

Patients who did not inspect their own results expressed that they were feeling fine and therefore had no interest in viewing their results.

Similarly, patients who were feeling fine or reported no impairments did not inspect the self-management tools, while others provided positive feedback on the self-management tools or the color-coding of questionnaire results.

> *I did not check the self-management advice because I am not feeling ill. Why would I check it?*
>
> *I like the design. The arrow in the results [direct link from the results to the self-management information] was very helpful.*
>
> *...liked the color coding [red/green] as it was simple and easy to understand.*

Four patients expressed the wish for a reminder (email or text message) before the next appointment to complete the questionnaires online. Two patients mentioned that they would like to be able to choose their password or username themselves (which the software currently does not allow). Three patients also explicitly reported a decline in motivation toward the end of the study, caused by a lack of sufficient feedback from physicians who did not discuss their PRO results during the consultations.

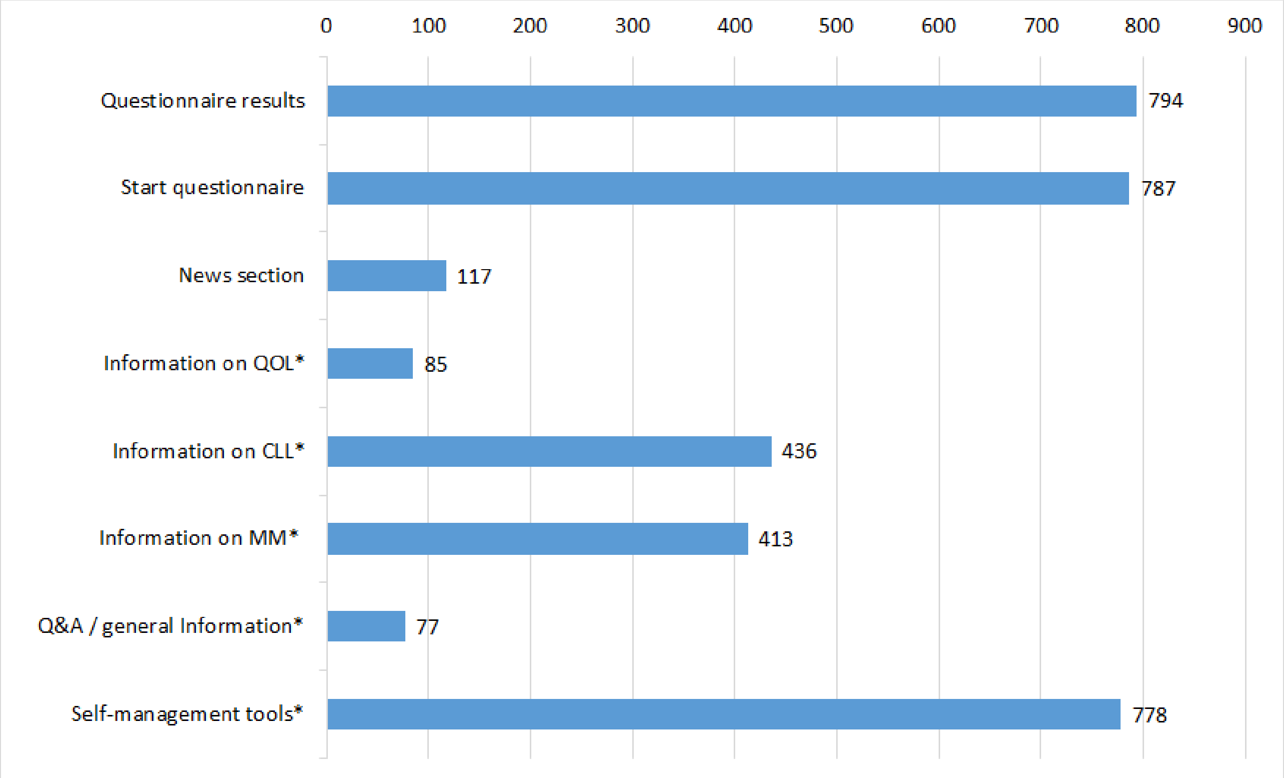
CHES Log Data

Over the study period, we registered a total of 796 sessions (ie, log-ins by patients). Of the patients who logged into the patient portal, 27% (28/102) logged in once, 18% (18/102) logged in twice, 9% (9/102) logged in 3 times, 15% (16/102) logged in 4 to 6 times, 14% (14/102) logged in 7 to 10 times, and 17% (17/102) logged in more than 10 times (total range of 1-57).

The mean duration of a session was 9.4 (median 6, range 1-90) minutes. It took patients on average 2.9 minutes to complete the EORTC QLQ-C30, 1.8 minutes to complete the EORTC QLQ-MY20, and 1.5 minutes to complete the EORTC QLQ-CLL17. This adds up to an average questionnaire completion time of 4.7 minutes for patients with MM and 4.4 minutes for patients with CLL.

During an average session, patients viewed 5.3 (median 4, range 1-28) different pages. Figure 4 shows how often specific components of the patient portal were viewed. A total of 3487 views were registered. The most frequently viewed self-management pages were those providing information on dyspnea (89 views), diarrhea (80 views), cognitive functioning (64 views), and emotional functioning (49 views). The least frequently viewed self-management pages were those providing information on obstipation (11 views), role functioning (12 views), social functioning (21 views), and pain (18 views). Multimedia Appendix 2 shows the view count on all self-management pages.

Figure 4. Page views in the patient portal by category (excluding home page). Categories marked with an asterisk combine multiple pages and subpages. QOL: quality of life; MM: multiple myeloma; CLL: chronic lymphocytic leukemia; Q&A: questions and answers.



Discussion

Principal Findings

In our study, we evaluated the use of the patient portal in patients with MM and CLL participating in routine electronic PRO monitoring. We found that in those patients who were eligible to use the patient portal, the majority (84%) were willing to use the portal. Only a few users reported difficulties logging in, but no users reported problems navigating the survey. However, we found that considerably fewer patients (37%) than initially included adopted a continued use of the portal across all 3 study time points. On average, patients spent 9.4 minutes in the portal per session.

Uptake of Home Monitoring

Our recruitment rate (ie, patients consenting to try the patient portal) is high compared to rates found in other studies. For example, other feasibility or usability studies report that between 21% and 64% [23-25] of patients were willing to use patient portals. Our high recruitment rate is most likely a consequence of the fact that patients were already participating in electronic PRO monitoring and so had already approved some form of monitoring. We would like to note the high inclusion rate for PRO monitoring at the hospital, which was, as reported previously, 94% of all MM patients treated in the department [18]. Our study emphasizes the possible synergy between assessments inside the hospital and home monitoring; patients who had already participated in electronic PRO monitoring inside the hospital were open to also using a patient portal. Moreover, the high acceptance of the patient portal might have been induced by the opportunity to get accustomed to ePROMs at the hospital with the help of a PRO facilitator before being introduced to the patient portal. We hypothesize that such a stepwise approach can reduce potential reservations about using electronic measures and contribute to user empowerment.

The adoption of home monitoring by a considerable proportion of patients also meant that those patients did not need to be assessed at the hospital. In this way, the patient portal reduced the administrative burden of assessing those patients' PROs at the hospital and allowed us to allocate those resources elsewhere.

Long-Term Use of the Portal

Over the course of the study, slightly less than half of the participants used the portal prior to their follow-up appointments. The main reason reported by patients was that they had forgotten to report their PRO data using the patient portal. Especially for patients with an infrequent appointment schedule (intervals between hospital visits of up to 12 months), forgetting to use the portal is, in fact, not surprising. This issue might be addressed by implementing an automated reminder (email or text message) as has been done with similar home monitoring systems [24,26]. For example, the AmbuFlex system in Denmark has, in the past, used different forms of reminders, including letters, emails, and text messages. They ultimately implemented communication via a national secure email program, which accounts for 93.2% of communication to patients and secures high completion rates [27].

Additionally, it is important to note that due to our routine monitoring approach, ePRO data from patients who did not use the portal prior to the follow-up appointments were not lost. Instead, those patients were invited to complete the ePROMs at the hospital as was done with other patients who did not use the portal.

While only a few of our patients directly mentioned a decline in motivation to complete ePROMs due to HCPs not picking up on the results during the consultations, this is in fact a frequent problem of PRO implementations in clinical settings [3,26,28,29]. For patients, there is little perceived benefit of completing the questionnaires if the results are not reviewed and discussed by HCPs. Instead, sharing their health status via questionnaires might even become burdensome. Therefore, it is important to engage HCPs with the concept of PROs and train them in the use of PRO results to prevent PRO data from becoming meaningless busywork that hinders clinical practice instead of enhancing it [28]. This requires the education and training of HCPs, which can be achieved by conducting specialized training programs [30]. A buy-in strategy may be used to increase HCP engagement with PROs, for example, via the adoption of the user interface to HCP needs and preferences, regular meetings, or analyses of PRO data in the registry upon request of HCPs. Another important approach is to make PRO data comprehensible and actionable for HCPs (eg, by providing advice on how to react to results) [26,31,32]. In our system, this is done, for example, by using thresholds for clinical importance for the EORTC QLQ-C30 [21] that facilitate the interpretation of patient PRO data.

Finally, the literature shows that in order to provide a sound theoretical basis for sustainable PRO solutions in routine care settings, an implementation science approach can be followed [33]. Implementation science can help identify barriers (which are often similar across contexts) and enablers (which often depend on the given hospitals' context) of patient-reported outcome measure (PROM) implementations.

User Patterns in the Portal

As has been found in other studies [13,17,25], the majority of our patients found the display of self-management advice linked with their results to be valuable. A recent review of electronic systems to measure PROs found that less than one-third (29%) of published electronic systems include features that provide tailored automated advice to patients, and less than half (41%) provide general educational information [17]. In a qualitative study evaluating another eHealth application, tailored feedback and advice was rated as appealing by most participants [34]. Participants valued the option of accessing information remotely between appointments and having a low threshold to receiving such information (compared to having to consult their treating physician or having to search for information) [34]. These are notions that echo the high number of views of pages with self-management advice we found in our study.

Nevertheless, despite the approval for self-management advice we found in our study, a recent clinical trial showed that, while patient portals can contribute to improved health-related quality of life [13], their measurable benefits on patient activation (knowledge, skills, and confidence for self-management) could not be shown in a diverse sample of cancer survivors [13]. However, previous evidence suggested that such effects may be more pronounced in newly diagnosed patients [25,35]. Another randomized controlled trial showed that weekly PRO monitoring along with the provision of tailored self-management advice, compared to usual care, significantly enhanced self-efficacy in patients with cancer [36].

Limitations

Our study was designed to be carried out during routine care with minimal interruption of the clinical workflow. Therefore, interview time points were integrated into the patients' hospital visits. This resulted in varying intervals between interviews (between 1 week and 1 year in a few cases), which may have influenced the results. However, we point out that these are real-world visitation schedules, and any application designed for routine care should be evaluated accordingly. Similarly, we included more males than females in our study. This reflects the epidemiology of MM and CLL, and the male/female ratio was comparable to data from the AMR.

Second, the CHES log data did not allow for detailed analysis of individual user page-view patterns but only for analysis of overall page views and duration and frequency of sessions. This means that patients logging on to the portal more frequently have a greater weight in the analysis of page views. Moreover, page view numbers have to be interpreted with care as they provide no information on whether the pages were actually read.

Another limitation is that the interviews were conducted by the authors (LN, PB, and JL), who introduced patients to the portal. Therefore, some patients might have been reluctant to voice criticism in the interviews, even though we actively encouraged patients to also report negative feedback.

Finally, we consider some sample-related limitations: We included only German-speaking patients in the patient portal. While the EORTC questionnaires are available in a large number of languages (and can be completed at the hospital [20]), translating and updating the content of the portal would require considerable resources. This results in a potential bias, and our findings may not be generalizable to patients with other first languages and limited German language proficiency. We are aware that this can create an imbalance in the provision of care, as this systematically excludes certain patient groups. In fact, patients not speaking the primary language of the country might profit most from receiving disease-specific information and from being able to report symptoms online in their native language. Moreover, we observed a selection bias due to patients' age (patients with MM were on average 11 years younger than the mean age of patients in the AMR), as older patients might be less proficient using the internet and were consequently not included in our study. While this is an important limitation to consider, its impact should decrease over time, as the population's internet capabilities have been steadily increasing [37] and should further increase in the years to come.

Future Steps

Following the insights gained in this study, we are currently in the process of updating our monitoring procedure and software. One important step is to more closely monitor patients' PROMs to swiftly identify deteriorations and act accordingly; we aim to have a trained onconurse monitor patient results and check on patients in case of deteriorations. If necessary, the nurse can schedule an earlier appointment or alert the treating HCP. Another planned improvement is to implement an email or text message reminder to improve PROM completion rates.

Conclusion

Our study shows that a patient portal enabling remote PRO data assessment can complement routine electronic PRO implementation at the hospital by reducing the burden of administration for the clinical team and offering an additional way for patients to engage with PROs. We found that the majority of patients were open to using the patient portal and interested in assessments from home. The low number of technical problems and absence of complaints demonstrate the general user-friendliness of our portal. While initial uptake was high, fewer patients adopted regular use of the portal prior to their appointments. To increase long-term participation rates, further motivational (eg, increasing HCP engagement with the PRO data) and technical (eg, email reminder) measures are needed.

Acknowledgments

We thank the patients who participated in this study for their valuable time and input, without which the study would not have been possible. We thank the Department of Internal Medicine V in Innsbruck and the Comprehensive Cancer Center Innsbruck for their support with the project. We thank Janny Sauter for her help in the collection of data.

Authors' Contributions

JL and PB wrote and edited the manuscript and collected and analyzed the data. BH, GR, JMG, LW, MS, RW, WWeyrer, and WWillenbacher conceptualized the study, supported the implementation of the patient portal, and edited the manuscript. LN conceptualized the study, supported the implementation of the patient portal, assisted in data collection, and edited the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

BH and GR have intellectual property rights to the software tool CHES. All remaining authors declare that they have no conflict of interest.

Multimedia Appendix 1
Examples of pages in the patient portal.
[PDF File (Adobe PDF File), 1495 KB - jmir_v23i6e26022_app1.pdf]

Multimedia Appendix 2
Total view count in the portal and specific view count for the self-management pages.
[PNG File , 104 KB - jmir_v23i6e26022_app2.png]

References

1. Patient-reported outcome measures: use in medical product development to support labeling claims. US Food and Drug Administration. 2009 Dec. URL: https://www.fda.gov/media/77832/download [accessed 2020-08-05]
2. 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]
3. Jensen RE, Snyder CF, Abernethy AP, Basch E, Potosky AL, Roberts AC, et al. Review of electronic patient-reported outcomes systems used in cancer clinical care. J Oncol Pract 2014 Jul;10(4):e215-e222. [doi: 10.1200/jop.2013.001067]
4. Digital Economy and Society Index (DESI) 2020: use of internet services. European Commission. URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=67075 [accessed 2020-09-03]
5. Detmar SB, Muller MJ, Schornagel JH, Wever LDV, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. JAMA 2002 Dec 18;288(23):3027-3034. [doi: 10.1001/jama.288.23.3027] [Medline: 12479768]
6. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724. [doi: 10.1200/jco.2004.06.078]
7. Basch E, Abernethy AP. Supporting clinical practice decisions with real-time patient-reported outcomes. J Clin Oncol 2011 Mar 10;29(8):954-956. [doi: 10.1200/JCO.2010.33.2668] [Medline: 21282536]
8. Takeuchi EE, Keding A, Awad N, Hofmann U, Campbell LJ, Selby PJ, et al. Impact of patient-reported outcomes in oncology: a longitudinal analysis of patient-physician communication. J Clin Oncol 2011 Jul 20;29(21):2910-2917. [doi: 10.1200/jco.2010.32.2453]
9. Barbera L, Sutradhar R, Seow H, Earle C, Howell D, Mittmann N, et al. Impact of standardized edmonton symptom assessment system use on emergency department visits and hospitalization: results of a population-based retrospective matched cohort analysis. J Oncol Pract 2020 Sep;16(9):e958-e965. [doi: 10.1200/jop.19.00660]
10. 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]
11. Lizée T, Basch E, Trémolières P, Voog E, Domont J, Peyraga G, et al. Cost-effectiveness of web-based patient-reported outcome surveillance in patients with lung cancer. J Thorac Oncol 2019 Jun;14(6):1012-1020 [FREE Full text] [doi: 10.1016/j.jtho.2019.02.005] [Medline: 30776447]
12. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Dec 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
13. van der Hout A, van Uden-Kraan CF, Holtmaat K, Jansen F, Lissenberg-Witte BI, Nieuwenhuijzen GAP, et al. Role of eHealth application Oncokompas in supporting self-management of symptoms and health-related quality of life in cancer survivors: a randomised, controlled trial. Lancet Oncol 2020 Jan;21(1):80-94. [doi: 10.1016/S1470-2045(19)30675-8] [Medline: 31838009]
14. Giesinger JM, Wintner LM, Zabernigg A, Gamper E, Oberguggenberger AS, Sztankay MJ, et al. Assessing quality of life on the day of chemotherapy administration underestimates patients' true symptom burden. BMC Cancer 2014 Oct 10;14:758 [FREE Full text] [doi: 10.1186/1471-2407-14-758] [Medline: 25305067]
15. Wintner LM, Sztankay M, Aaronson N, Bottomley A, Giesinger JM, Groenvold M, EORTC Quality of Life Group. The use of EORTC measures in daily clinical practice: a synopsis of a newly developed manual. Eur J Cancer 2016 Nov;68:73-81 [FREE Full text] [doi: 10.1016/j.ejca.2016.08.024] [Medline: 27721057]
16. LeRouge C, Austin E, Lee J, Segal C, Sangameswaran S, Hartzler A, et al. ePROs in Clinical Care: Guidelines and Tools for Health Systems. Seattle: CERTAIN, University of Washington; 2020.
17. Warrington L, Absolom K, Conner M, Kellar I, Clayton B, Ayres M, et al. Electronic systems for patients to report and manage side effects of cancer treatment: systematic review. J Med Internet Res 2019 Jan 24;21(1):e10875 [FREE Full text] [doi: 10.2196/10875] [Medline: 30679145]
18. Sztankay M, Neppl L, Wintner LM, Loth FL, Willenbacher W, Weger R, et al. Complementing clinical cancer registry data with patient reported outcomes: a feasibility study on routine electronic patient-reported outcome assessment for the

Austrian Myelome Registry. Eur J Cancer Care (Engl) 2019 Nov;28(6):e13154 [FREE Full text] [doi: 10.1111/ecc.13154] [Medline: 31465136]

19. Holzner B, Giesinger JM, Pinggera J, Zugal S, Schöpf F, Oberguggenberger AS, et al. The Computer-based Health Evaluation Software (CHES): a software for electronic patient-reported outcome monitoring. BMC Med Inform Decis Mak 2012;12:126 [FREE Full text] [doi: 10.1186/1472-6947-12-126] [Medline: 23140270]
20. Lehmann J, Wintner LM, Sztankay M, Willenbacher W, Weger R, Weyrer W, et al. Patient-reported outcomes and psycho-oncological screening in hematology: a practical example of routine electronic monitoring. Mag Eur Med Oncol 2020 Jul 03;13(3):285-293. [doi: 10.1007/s12254-020-00628-7]
21. Giesinger JM, Loth FLC, Aaronson NK, Arraras JI, Caocci G, Efficace F, EORTC Quality of Life Group. Thresholds for clinical importance were established to improve interpretation of the EORTC QLQ-C30 in clinical practice and research. J Clin Epidemiol 2020 Feb;118:1-8 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.10.003] [Medline: 31639445]
22. Efficace F, Collins GS, Cottone F, Giesinger JM, Sommer K, Anota A, et al. Patient-reported outcomes as independent prognostic factors for survival in oncology: systematic review and meta-analysis. Value Health 2021 Feb;24(2):250-267. [doi: 10.1016/j.jval.2020.10.017] [Medline: 33518032]
23. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. eHealth for breast cancer survivors: use, feasibility and impact of an interactive portal. JMIR Cancer 2016 May 10;2(1):e3 [FREE Full text] [doi: 10.2196/cancer.5456] [Medline: 28410178]
24. Girgis A, Durcinoska I, Levesque JV, Gerges M, Sandell T, Arnold A, PROMPT-Care Program Group. eHealth system for collecting and utilizing Patient Reported Outcome Measures for Personalized Treatment and Care (PROMPT-Care) among cancer patients: mixed methods approach to evaluate feasibility and acceptability. J Med Internet Res 2017 Dec 02;19(10):e330 [FREE Full text] [doi: 10.2196/jmir.8360] [Medline: 28970188]
25. Duman-Lubberding S, van Uden-Kraan CF, Jansen F, Witte BI, van der Velden LA, Lacko M, et al. Feasibility of an eHealth application "OncoKompas" to improve personalized survivorship cancer care. Support Care Cancer 2016 May;24(5):2163-2171 [FREE Full text] [doi: 10.1007/s00520-015-3004-2] [Medline: 26563178]
26. Wu AW, White SM, Blackford AL, Wolff AC, Carducci MA, Herman JM, PatientViewpoint Scientific Advisory Board. Improving an electronic system for measuring PROs in routine oncology practice. J Cancer Surviv 2016 Jun;10(3):573-582 [FREE Full text] [doi: 10.1007/s11764-015-0503-6] [Medline: 26644190]
27. Hjollund NHI. Fifteen years' use of patient-reported outcome measures at the group and patient levels: trend analysis. J Med Internet Res 2019 Sep 30;21(9):e15856 [FREE Full text] [doi: 10.2196/15856] [Medline: 31573943]
28. Thestrup Hansen S, Kjerholt M, Friis Christensen S, Hølge-Hazelton B, Brodersen J. Haematologists' experiences implementing patient reported outcome measures (PROMs) in an outpatient clinic: a qualitative study for applied practice. J Patient Rep Outcomes 2019 Dec 28;3(1):74 [FREE Full text] [doi: 10.1186/s41687-019-0166-6] [Medline: 31884569]
29. Taarnhøj GA, Lindberg H, Dohn LH, Omland LH, Hjøllund NH, Johansen C, et al. Electronic reporting of patient-reported outcomes in a fragile and comorbid population during cancer therapy: a feasibility study. Health Qual Life Outcomes 2020 Jul 11;18(1):225 [FREE Full text] [doi: 10.1186/s12955-020-01480-3] [Medline: 32653005]
30. Santana MJ, Haverman L, Absolom K, Takeuchi E, Feeny D, Grootenhuis M, et al. Training clinicians in how to use patient-reported outcome measures in routine clinical practice. Qual Life Res 2015 Jan 15;24(7):1707-1718. [doi: 10.1007/s11136-014-0903-5]
31. Snyder CF, Blackford AL, Wolff AC, Carducci MA, Herman JM, Wu AW, PatientViewpoint Scientific Advisory Board. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. Psychooncology 2013 Apr;22(4):895-901 [FREE Full text] [doi: 10.1002/pon.3087] [Medline: 22544513]
32. Hughes EF, Wu AW, Carducci MA, Snyder CF. What can I do? Recommendations for responding to issues identified by patient-reported outcomes assessments used in clinical practice. J Support Oncol 2012;10(4):143-148 [FREE Full text] [doi: 10.1016/j.suponc.2012.02.002] [Medline: 22609239]
33. Stover AM, Haverman L, van Oers HA, Greenhalgh J, Potter CM, ISOQOL PROMs/PREMs in Clinical Practice Implementation Science Work Group. Using an implementation science approach to implement and evaluate patient-reported outcome measures (PROM) initiatives in routine care settings. Qual Life Res 2020 Jul 10:1. [doi: 10.1007/s11136-020-02564-9] [Medline: 32651805]
34. Lubberding S, van Uden-Kraan CF, Te Velde EA, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors. J Clin Nurs 2015 May;24(9-10):1367-1379. [doi: 10.1111/jocn.12753] [Medline: 25677218]
35. Girgis A. The role of self-management in cancer survivorship care. Lancet Oncol 2020 Jan;21(1):8-9. [doi: 10.1016/s1470-2045(19)30715-6]
36. Absolom K, Warrington L, Hudson E, Hewison J, Morris C, Holch P, et al. Phase III randomized controlled trial of eRAPID: eHealth intervention during chemotherapy. J Clin Oncol 2021 Mar 01;39(7):734-747. [doi: 10.1200/JCO.20.02015] [Medline: 33417506]
37. Austrian Internet Monitor. URL: https://de.statista.com/ [accessed 2021-02-08]

Abbreviations

AMR: Austrian Myeloma Registry
CHES: Computer-Based health Evaluation System
CLL: chronic lymphocytic leukemia
EORTC QLQ–C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire
ePROM: electronic patient-reported outcome measure
HCP: health care professional
MM: multiple myeloma
PRO: patient-reported outcome
PROM: patient-reported outcome measure

Edited by R Kukafka; submitted 25.11.20; peer-reviewed by L Schougaard, T Sandell, S Oerlemans; comments to author 20.01.21; revised version received 08.02.21; accepted 30.04.21; published 08.06.21.

Please cite as:
Lehmann J, Buhl P, Giesinger JM, Wintner LM, Sztankay M, Neppl L, Willenbacher W, Weger R, Weyrer W, Rumpold G, Holzner B
Using the Computer-based Health Evaluation System (CHES) to Support Self-management of Symptoms and Functional Health: Evaluation of Hematological Patient Use of a Web-Based Patient Portal
J Med Internet Res 2021;23(6):e26022
URL: https://www.jmir.org/2021/6/e26022
doi:10.2196/26022
PMID:34100765

Original Paper

Reliable Prediction Models Based on Enriched Data for Identifying the Mode of Childbirth by Using Machine Learning Methods: Development Study

Zahid Ullah[1*], PhD; Farrukh Saleem[1*], PhD; Mona Jamjoom[2*], PhD; Bahjat Fakieh[1*], PhD

[1]Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
[2]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia
[*]all authors contributed equally

Corresponding Author:
Mona Jamjoom, PhD
Department of Computer Sciences, College of Computer and Information Sciences
Princess Nourah Bint Abdulrahman University
PO Box 84428
Riyadh, 11671
Saudi Arabia
Phone: 966 505273052
Email: mmjamjoom@pnu.edu.sa

Abstract

Background: The use of artificial intelligence has revolutionized every area of life such as business and trade, social and electronic media, education and learning, manufacturing industries, medicine and sciences, and every other sector. The new reforms and advanced technologies of artificial intelligence have enabled data analysts to transmute raw data generated by these sectors into meaningful insights for an effective decision-making process. Health care is one of the integral sectors where a large amount of data is generated daily, and making effective decisions based on these data is therefore a challenge. In this study, cases related to childbirth either by the traditional method of vaginal delivery or cesarean delivery were investigated. Cesarean delivery is performed to save both the mother and the fetus when complications related to vaginal birth arise.

Objective: The aim of this study was to develop reliable prediction models for a maternity care decision support system to predict the mode of delivery before childbirth.

Methods: This study was conducted in 2 parts for identifying the mode of childbirth: first, the existing data set was enriched and second, previous medical records about the mode of delivery were investigated using machine learning algorithms and by extracting meaningful insights from unseen cases. Several prediction models were trained to achieve this objective, such as decision tree, random forest, AdaBoostM1, bagging, and k-nearest neighbor, based on original and enriched data sets.

Results: The prediction models based on enriched data performed well in terms of accuracy, sensitivity, specificity, F-measure, and receiver operating characteristic curves in the outcomes. Specifically, the accuracy of k-nearest neighbor was 84.38%, that of bagging was 83.75%, that of random forest was 83.13%, that of decision tree was 81.25%, and that of AdaBoostM1 was 80.63%. Enrichment of the data set had a good impact on improving the accuracy of the prediction process, which supports maternity care practitioners in making decisions in critical cases.

Conclusions: Our study shows that enriching the data set improves the accuracy of the prediction process, thereby supporting maternity care practitioners in making informed decisions in critical cases. The enriched data set used in this study yields good results, but this data set can become even better if the records are increased with real clinical data.

(J Med Internet Res 2021;23(6):e28856) doi:10.2196/28856

KEYWORDS

machine learning; prediction model; health care; cesarean; delivery; decision making

Introduction

Background

Machine learning is increasingly prevalent in and vital to health care industries in terms of predicting and identifying quality treatments for patients and enhancing other health care services. Therefore, machine learning techniques are used for extracting knowledge from huge and complex data sets in an organized form so that it can be used for making effective decisions. According to Sana et al [1], machine learning techniques provide diagnosis and analytical amenities in several medical fields and their applications in clinical factors and analytics such as disease prediction, decision making based on extracted medical knowledge, and serving in patient management. Moreover, with the increasing amount of available data, machine learning techniques have significant benefits as prediction tools in health care [2] that sometimes provide surprising prediction models that help in clinical counseling [3]. These tools are fundamental to biomedical research and are utilized as an integral part of the clinical decision-making process [4].

Child delivery can be performed through several methods in hospitals, but the most common methods are either traditional vaginal birth or cesarean (c-section), while vacuum extractions and obstetric pincers can be used during complications in vaginal deliveries [5]. There are several assumptions pertaining to the mode of delivery, but it is still challenging to predict the type of childbirth accurately [6,7]. C-section is a technique used in maternity care for delivering children by performing a surgical incision to the woman’s abdomen and uterus [8], which normally takes place when complications arise related to the mother or a child in a normal delivery [9]. The possible complications of c-section for mothers are infections, excessive bleeding that could cause anemia, and reaction to anesthesia; therefore, maternal death rates with c-sections are higher than that for vaginal deliveries [8]. However, a c-section could be necessary to save the lives of both the mother and the child if the baby is located in a wrong position in the womb, the head of the baby is larger than the birth canal, the direction of the baby is reversed, or the mother has a c-section history or even heart-related diseases [10]. Molina et al [11] further explained that c-sections are lifesaving for obstructed labor and any other obstruction in the delivery process for decreasing baby and mother mortality, but the risk of complications and overuse can harm both mothers and babies. Every mode of delivery has its pros and cons, but selecting the wrong type may lead to a variety of risks such as baby cessation, excessive bleeding, baby breathing problems, and other similar issues [7].

The rate of c-sections is higher than the rate of normal deliveries, especially in high-income countries, where in 2012, around 23 million deliveries were conducted by c-section worldwide [11]. Prema and Pushpalatha [8] indicated that the highest rate of c-section was 29.1% in November 2005, while nearly one-third of the deliveries were conducted using c-section in 2015 [12] as reported by the Centers for Disease Control and Prevention [13]. In the United States, the c-section rate significantly increased to 60% from 1996 to 2009, and the c-section rate was 32% of all deliveries in 2007 [12]. Li et al [14] reported that in China, 46.2% of the 14,541 deliveries across 3 provinces in different hospitals were conducted by c-section in the years 2007 and 2008. Similarly, in Pakistan, around half of the total deliveries are conducted at home, but a high number of c-sections are conducted at hospitals [9]. Fergus et al [4] argued that overinterpretations increase the numbers of c-section, even if there are no specific risks involved in the normal deliveries. It is difficult to know the optimal level of the c-section rate because although the World Health Organization advocates that national rates do not exceed by 10 to 15 c-sections per 100 births, the rates of c-sections are noticeably higher [11].

Related Work

Studies related to identifying the mode of childbirth were found in different databases such as Google Scholar, Science Direct, IEEE explorer, Wiley, ResearchGate, and other data sources. The major keywords used in the browsers were phrases such as cesarean sections using machine learning, c-sections using machine learning, machine learning in maternity care, AI in maternity care, etc. C-section is the most commonly increasing mode of delivery worldwide, and areas of concerns such as ideal c-section rate, safety, and cost are still under debate [15]. Moreover, many researchers have investigated different features to determine the main causes for cesarean delivery and have built a prediction model based on these features. Some main causes are related to the medical and obstetric history of the mother [5,6,8,15-17]. The study of Lee and Gay [18] found that sleep disturbance and fatigue in late pregnancy lead to greater chances of delivery by c-section. Others analyzed the socioeconomic or sociodemographic features [1,19] and some have determined the main causes to be the region and level of medical services afforded [9].

Wollmann et al [3] attempted to predict the chances of normal births after a c-section. In this regard, they collected data of women with one previous birth in Sweden during 2008-2014 and built 3 machine learning models and 1 regression model. They concluded that the majority of the women with a history of c-sections could still successfully deliver a baby in the normal way. Similarly, Prema and Pushpalatha [8] investigated the main causes of cesarean delivery based on the extracted features. Several machine learning models were trained on a data set collected from a pregnancy risk assessment survey. Their models have predicted c-sections with 96% accuracy for women who had a history of c-sections compared to 89% accuracy for women who had no previous c-section. Khan et al [10] presented a study to predict whether c-section is compulsory along with increased safety for both mother and child during and after delivery. They trained 3 ensemble models and found the highest accuracy model of 87.66%. They also found that for predicting the target mode of delivery, several features such as previous c-sections, amniotic fluid, fetal intrapartum pH, and preinduction should be considered. Sana et al [1] figured out the socioeconomic features that cause cesarean delivery. They trained decision tree (DT) and artificial neural network models to predict the mode of delivery in which artificial neural networks showed a high accuracy of 82%. Abbas et al [9] believed that c-section causes can be influenced by regions and therefore, they selected a region with a limited health care infrastructure. They trained several models based on 23 features in order to predict the mode of delivery, and the highest accuracy model was 91.8%. They also concluded that the maternal age and the previous mode of delivery considerably influenced the mode of the next delivery.

Ricciardi et al [20] adopted classification methods DT, random forest (RF), AdaBoostM1, gradient boosting, and DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples) for predicting patients' mode of delivery. They applied these methods to a data set of 370 records collected from public and private hospitals from the years 2000 to 2009. RF outperformed with 91.1% accuracy, 90% sensitivity, and >96% ROC. In the study of Improta et al [2], the 4 classification methods, namely, DT, RF, AdaBoost, and gradient boosting were trained on a cardiotocographic data set for identifying the mode of delivery, in which RF showed the highest performance with 87.6% accuracy, 87.9% precision, and 93% ROC. In a study conducted by Saleem et al [21] for classifying the mode of delivery using 4 machine learning methods, the AdaBoost model showed the highest accuracy of 91.8%, sensitivity of 95.5%, and specificity of 98%. Of the 4 classification algorithms used by Pereira et al [5] to predict the mode of delivery, DT outperformed with accuracy of 84%, sensitivity of 88%, and specificity of 80%. A DT method was adopted by Soleimanian et al [15] to investigate the mode of delivery in a data set of 80 patients and they found an accuracy of 86.25%. In the study of Fergus et al [4], ensemble methods were used for classifying the mode of delivery by using a cardiotocographic tracer in which all 3 methods showed promising results of 87%, 90%, and 96% for sensitivity, specificity, and ROC, respectively. Moreover, Fergus et al [22] established that machine learning with fetal heart rate signals significantly improved the efficacy of detecting the mode of delivery compared to obstetrician and midwife predictions and other systems. Their results showed 94%, 91%, and 99% sensitivity, specificity, and ROC, respectively.

Objective of This Study

This study aims to provide prediction models for identifying the mode of childbirth based on antenatal signs and symptoms by using machine learning techniques. To achieve the objectives of this study, the data set was first enriched with additional cases using the Synthetic Minority Oversampling Technique (SMOTE) [23]. Second, several prediction models were trained and tested on original and enriched data sets. A cross-validation of 10 folds was used for evaluating the performance of the models. In the outcomes, the enriched data set showed better performance in terms of accuracy, sensitivity, specificity, F-measure, and receiver operating characteristic (ROC) compared to the original data set. These findings encourage the applications of these models for maternity care decision support systems to predict the mode of delivery before birth.

Methods

Software Used

The data synthesis and analysis in terms of classifications and predicting the mode of delivery were performed using Weka software (University of Waikato, New Zealand) [24]. Weka has many machine learning algorithms that are useful for training data sets and then testing them on unseen cases to predict target values [25,26].

Data Collection

The data set used in this study was harvested from the study of Soleimanian et al [15] and is publicly available in the University of California, Irvine machine learning repository [12,27]. The data set contains 5 features, namely, age, delivery_number, delivery_time, blood_of_pressure, and heart_problem, while cesarean is a class attribute to label whether the delivery was performed by c-section. In the data set, each attribute shows different values, such as age ranges from 22 years to 38 years, delivery_number shows the number of births from 1 to 4, delivery_time shows 3 different statuses that are premature, timely, or latecomer, blood_of_pressure also shows 3 different statuses that are low, normal, and high, the heart_problem is categorized as either yes or no, and the last attribute (cesarean) is categorized as to whether the birth was by c-section or not. This data set contains 80 records of pregnant women and information about whether delivery was conducted by c-section or normal birth.

Data Enrichment

The data set used in this study originally contained 80 records, of which 46 records were normal vaginal deliveries while the remaining 34 were c-sections. According to the criteria of Vapnik [28], the total number of records was insufficient for predictive purposes [25]. Therefore, the data set needed more records to ensure that the prediction models are reliable and trustworthy. For this reason, the existing data set was enriched with more records using the standard method of SMOTE [23]. SMOTE is a popular method of machine learning used for oversampling [29] in which the minority class in a data set is generated by a synthetic example in the feature area based on the selected k-nearest neighbor (k-NN) from the minority class [21]. This practice has been adopted in several biomedical studies [4,30-36]. Mohammed et al [34] used the SMOTE method for enriching the minority class and concluded that oversampling has a positive impact on the prediction models. Similarly, Ramezankhani et al [32] adopted the SMOTE method for increasing the samples in the minority class in the original data set with various percentages (ie, 100%, 200%....,700%), which resulted in increased sensitivity of the different classifiers used. Another study of Hussain et al [37] used the SMOTE method and compared the results with the original data, concluding that the prediction models' performance after oversampling was enhanced compared to the original data. According to Ebenuwa [38] and Frank [39], SMOTE can be used for increasing the size of a data set. This study used the SMOTE method for enriching the samples in both classes with 100%; therefore, the total number of records after oversampling increased to 160 while the ratio between the 2 classes remained the same as in the original data set. At the current stage, the enriched data set was sufficient for reliable prediction.

Prediction Models

The ability to gain meaningful insight from the available unstructured and unorganized data and to utilize it as an integral part of a business decision support system is an art. There are several technologies available that work in this domain for structuring and organizing the historical data for predicting new patterns for the unseen scenarios, including machine learning. Similarly, these techniques are widely used in the health care industry, where prediction models have evolved with clinical practice in every medical field. In the literature, several studies have attempted to classify the types of childbirth from different perspectives by using machine learning models. The most widely used classifiers for predictions are DT, RF, AdaBoost, support vector machine, k-NN, Naïve Bayes, and several other techniques. This study has utilized 5 machine learning classifiers for developing prediction models that can help health care practitioners in deciding the favorable mode of delivery, primarily based on the mother's history and condition. A brief discussion of these classifiers is presented below.

DT

DT is a nonparametric supervised learning technique used for both classification and regression and it uses large and complicated data sets to explore features and mined patterns that are vital to discrimination and predictive modeling. In this technique, the large data sample is divided into training and testing data sets, and based on the training data set, building a DT model and a testing data set in order to decide on the suitable tree size required to attain an optimal final model is performed [40].

RF

RF is an ensemble technique used for classification or regression that utilizes the input data and constructs multitude of DTs at the training time and outputs the class (classification) or the prediction mean (regression) of an individual tree [41]. In this technique, each DT is randomized using a bootstrap resampling method with random feature selection, and the classification is performed based on the voting of various randomized DTs on the final outcome [4]. Furthermore, the optimum split is computed using various feature sets and lingers until the tree is completely grown without pruning. This process is iterated for all trees in the forest by using different bootstraps of data, and the classifications of new samples are therefore based on the majority of votes cast [4].

AdaBoost

AdaBoost is an ensemble technique of linear member classifiers that is constructed to enhance the efficiency of the binary classifiers. In this technique, the weak learning models with better accuracy can be boosted to develop a strong prediction model. AdaBoost is an iterative-based technique where each iteration detects the misclassified data points and increases the weights of the correct points to increase the chance of the next classifiers getting them right. Moreover, in this method, the instances are moved from the iterative samples of the training data to the subsequent data set, and the classifiers are combined based on the weighted majority of votes [10,42].

Bagging

Bootstrap aggregation (or bagging) is an ensemble technique used for classification or regression. In bagging, a repeated sample is made from a training set by using simple random sampling with replacement, and for each bootstrap sample, a weak classifier is trained. These classifiers are then utilized for predicting class labels on testing data, and the class that obtains the majority of the votes wins [43].

k-NN

k-NN is a supervised learning technique that takes a data set in which the data points are labeled with different classes and uses them for learning to label the new points. The labeling of new points is based on the closest of its neighbors' labels and the majority of votes cast; therefore, the labels of the nearest neighbors are the labels of the new points. In k-NN, k is the criteria number of checking the nearest neighbors [12,44].

Performance Evaluation Method

There are several methods for evaluating the performance of prediction models such as using the whole data set as a training set, providing a separate test set, cross-validation, and percentage split, of which cross-validation is regarded as the most reliable method [45]. In this study, each prediction model built was evaluated using cross-validations of 10 folds [46]. In 10-fold cross-validation, the training set is divided into 10 subsets, and each subset is used once in the testing phase [47]. Amin and Ali [12] and Soleimanian et al [15] trained their models by using the whole data set as training data, but this method was not recommended in several other studies such as those of Mitchell [48], Smith and Frank [45], and Brownlee [49] because machine learning methods learn the training data and can predict them easily. As explained by Mitchell [48], utilizing the entire data set for training and testing purposes at the same time may produce unrealistic outcomes that are extremely positive and prone to overfitting. As further explained by Smith and Frank [45], the results achieved using training data as test data give rise to resubstituting errors, which are typically unjustifiably optimistic for predicting the performance of a model with future unseen data. Moreover, a training set for a model evaluation can be useful if one is more interested in a descriptive rather than a predictive model [49]. This is usually the challenge of machine learning: to predict unseen cases that have not been trained. To the contrary, cross-validation is regarded as the most profound and reliable method for model evaluation in machine learning when all data exist in 1 set [45]. Furthermore, in cross-validation, the test set contains unseen cases that are unknown to the model during the training phase, which can help reliable assessment of a classifier's performance [50] because cross-validation helps render generalization errors and variance [51]. As further explained by Schaffer [52] cross-validation can be used to choose a classifier in case of lack of pertinent domain-specific knowledge. In short, cross-validation provides practical estimation because a model is predicting actual results that may have been unknown to the model in the training process.

Results

This study has applied the selected prediction models to both the original data set (80 cases) and the enriched data set (160 cases). The performance of each model was evaluated using cross-validations of 10 folds [46]. In 10-fold cross-validation, the training set is divided into 10 subsets, and each subset is used once in the training phase [47]. In the implementation phase, when the required parameters are set for testing the model, a confusion matrix is calculated for each classifier run. Specifically, the confusion matrix provides 4 important values that are computed based on the correctly and incorrectly classified instances of a data set. These values are commonly known as true positive, true negative, false positive, and false negative. This matrix is the basis for calculating important measures such as model performance, model accuracy, sensitivity, specificity, and F-measure. All these measures are calculated using different equations. For example, the accuracy of a model is calculated using the following equation:

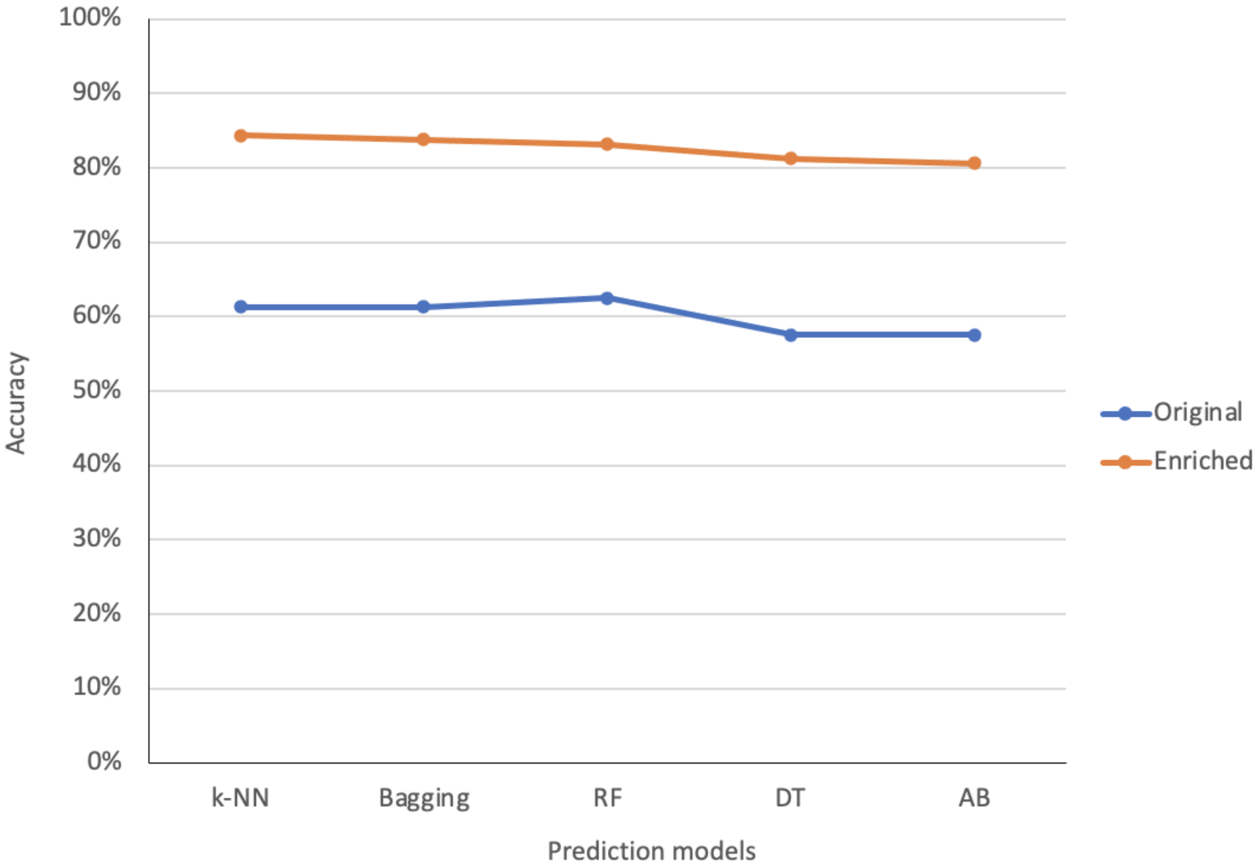
where FP=false positive, FN=false negative, TP=true positive, and TN=true negative.

The accuracies of the different models before and after enriching the data set for identifying the mode of delivery were evaluated using equation 1, and the outcomes are depicted in Table 1 and Figure 1.

Table 1. Accuracies and kappa values of models before and after data enrichment.

Method	Original data set		Enriched data set	
	Accuracy (%)	Kappa value	Accuracy (%)	Kappa value
k-Nearest neighbor	61.25	0.228	84.38	0.685
Bagging	61.25	0.192	83.75	0.664
Random forest	62.50	0.215	83.13	0.654
Decision tree	57.50	0.181	81.25	0.612
AdaBoost	57.50	0.124	80.63	0.603

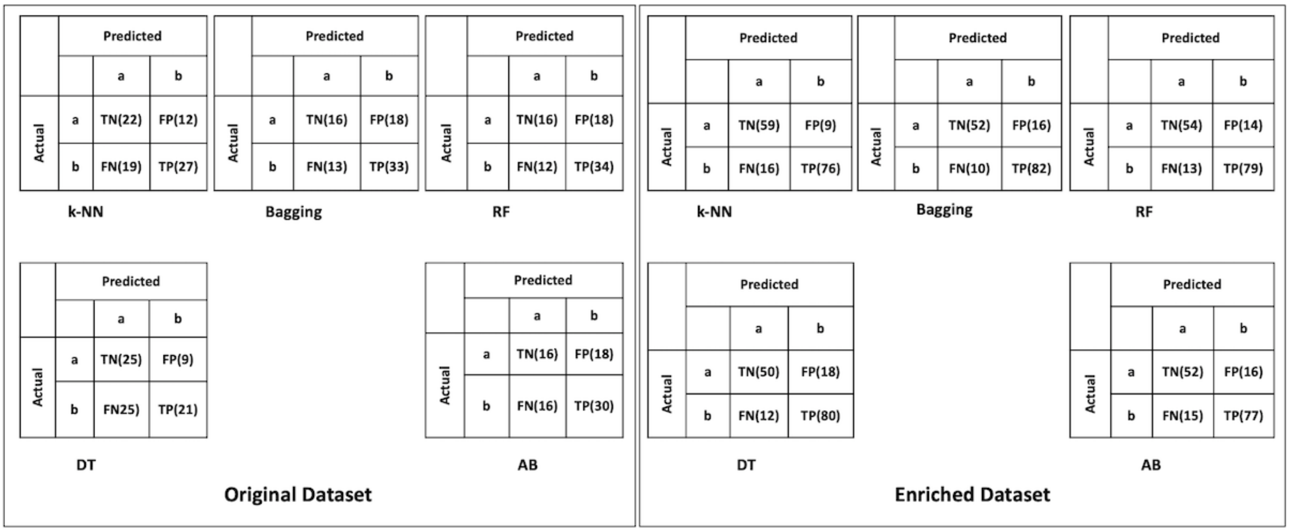
Figure 1. Comparison of the accuracy between models before and after data enrichment. AB: AdaBoost; DT: decision tree; k-NN: k-nearest neighbor; RF: random forest.



According to Table 1 and Figure 1, the performances of all the models in terms of accuracy were very low when they were trained with the original data set; however, accuracy was tremendously improved when the models were trained with the updated enriched data set, whose improvement reached approximately 20%-23%. In the original data set, RF showed the highest accuracy of 62.50%, which was far lower in performance than the lower model trained with the enriched data. Moreover, for the models trained with the enriched data set, k-NN showed the highest accuracy of 84.38%, while bagging, RF, DT, and AdaBoost showed accuracies of 83.75%, 83.13%, 81.25%, and 80.63%, respectively.

Kappa values or kappa statistics [53] is a measure that compares the observed accuracy with the expected accuracy (random chance) and is the appropriate method when 2 or more independent classifiers are analyzing the same case [9]. There are different thresholds ranges for the kappa values [54]; however, in machine learning, when investigating an unseen scenario, a kappa value higher than 0.40 might be considered exceptional [55]. According to Table 1, the kappa values of k-NN, bagging, RF, DT, and AdaBoost in the original data set are lower than the threshold, but in the enriched data set are 0.685, 0.664, 0.654, 0.612, and 0.603, respectively, surpassing the threshold value. Moreover, Figure 2 shows the confusion matrix of the models used in this study, where “a” represents 0 class while “b” represents 1 class in the data set. Furthermore, the accuracies of the models for identifying the mode of delivery were also measured using recall, precision, and F-measure. These are the important measures computed based on the values of the confusion matrix. Recall, which is also referred to as sensitivity, is the proportion of the real positive values that are correctly classified as positive, while precision, which is referred to as predictive positive value or confidence [56] or specificity [57], is the proportion of the predicted positive values that are correctly real positives [56]. Similarly, F-measure [58] is the hormonic mean of precision and recall [59]. Table 2 and Table 3 exhibit the values of recall, precision, and F-measures for all models trained before and after data enrichment, respectively.

Figure 2. Confusion matrix of applied models before and after data enrichment. AB: AdaBoost; DT: decision tree; FN: false negative; FP: false positive; k-NN: k-nearest neighbor; RF: random forest; TN: true negative; TP: true positive.



Original Dataset

		Predicted	
		a	b
Actual	a	TN(22)	FP(12)
	b	FN(19)	TP(27)

k-NN

		Predicted	
		a	b
Actual	a	TN(16)	FP(18)
	b	FN(13)	TP(33)

Bagging

		Predicted	
		a	b
Actual	a	TN(16)	FP(18)
	b	FN(12)	TP(34)

RF

		Predicted	
		a	b
Actual	a	TN(25)	FP(9)
	b	FN25)	TP(21)

DT

		Predicted	
		a	b
Actual	a	TN(16)	FP(18)
	b	FN(16)	TP(30)

AB

Enriched Dataset

		Predicted	
		a	b
Actual	a	TN(59)	FP(9)
	b	FN(16)	TP(76)

k-NN

		Predicted	
		a	b
Actual	a	TN(52)	FP(16)
	b	FN(10)	TP(82)

Bagging

		Predicted	
		a	b
Actual	a	TN(54)	FP(14)
	b	FN(13)	TP(79)

RF

		Predicted	
		a	b
Actual	a	TN(50)	FP(18)
	b	FN(12)	TP(80)

DT

		Predicted	
		a	b
Actual	a	TN(52)	FP(16)
	b	FN(15)	TP(77)

AB

Table 2. Precision, recall, and F-measure of the models trained with original data.

Method	Cesarean section			Normal delivery		
	Precision	Recall	F-measure	Precision	Recall	F-measure
k-Nearest neighbor	0.692	0.587	0.635	0.537	0.647	0.587
Bagging	0.647	0.717	0.680	0.552	0.471	0.508
Random forest	0.654	0.739	0.694	0.571	0.471	0.516
Decision tree	0.700	0.457	0.553	0.500	0.735	0.595
AdaBoost	0.625	0.652	0.638	0.500	0.471	0.485

Table 3. Precision, recall, and F-measure of the models trained with enriched data.

Method	Cesarean section			Normal delivery		
	Precision	Recall	F-measure	Precision	Recall	F-measure
k-Nearest neighbor	0.894	0.826	0.859	0.787	0.868	0.825
Bagging	0.837	0.891	0.863	0.839	0.765	0.800
Random forest	0.849	0.859	0.854	0.806	0.794	0.800
Decision tree	0.816	0.870	0.842	0.806	0.735	0.769
AdaBoost	0.828	0.837	0.832	0.776	0.765	0.770

There is a clear difference between the values of all measures in Table 2 and Table 3 due to the feeding of additional records into the data set for data enrichment. The models' performance based on the enriched data set has shown the values of precision, recall, and F-measure above 80% accuracy, except for a few values in Table 3. This is empirical evidence that populating the data set with additional records can increase the performance of the prediction models. Hence, Table 3 supports that these models can be used for maternity care decision making in identifying the mode of delivery before birth. Similarly, the models were analyzed using ROC curve evaluation [60]. ROC curves are highly useful for establishing the classifiers and envisioning their performance and are commonly used in health care decision making [61] because it visualizes the entire scenario of trade-off between recall and (1-specificity) across a set of cutoff points and is considered an effectual measure of inherent validity of a diagnostic test [62]. Moreover, as discussed in a previous study [9], ROC curves provide the percentage between precision and recall in which higher values of precision represent a low false-positive rate, which means that the classifier returns an accurate outcome, and the high values of recall showing a low false-negative rate, which means that the classifier returns positive outcomes. Figure 3 and Figure 4 show the ROC curves of all classifiers used for predicting the mode of delivery based on before and after data enrichment, respectively. The ROC curve has several advantages over single values of precision and recall in which one of its important benefits is that 2 or more diagnostic tests can be graphically compared at the same time in 1 graph [62]. Moreover, a curve that is nearer to the left upper corner shows the best accuracy of a classifier, while a curve closer to the lower right corner shows the worst [63]. In Figure 4, the curves closer to the left upper corner provide solid evidence, indicating that the accuracies of the classifiers used in the models based on enriched data are high. Therefore, these models are reliable and can be used for predicting the mode of delivery in the antenatal stage and can also be a part of the maternity care decision support system. Figure 3 shows ROC curves as middle lines, which are far away from the left upper corner compared to Figure 4; thus, Figure 4 is significantly more reliable.

Figure 3. Receiver operating characteristic curves of all classifiers based on original data. AB: AdaBoost; DT: decision tree; k-NN: k-nearest neighbor; RF: random forest.

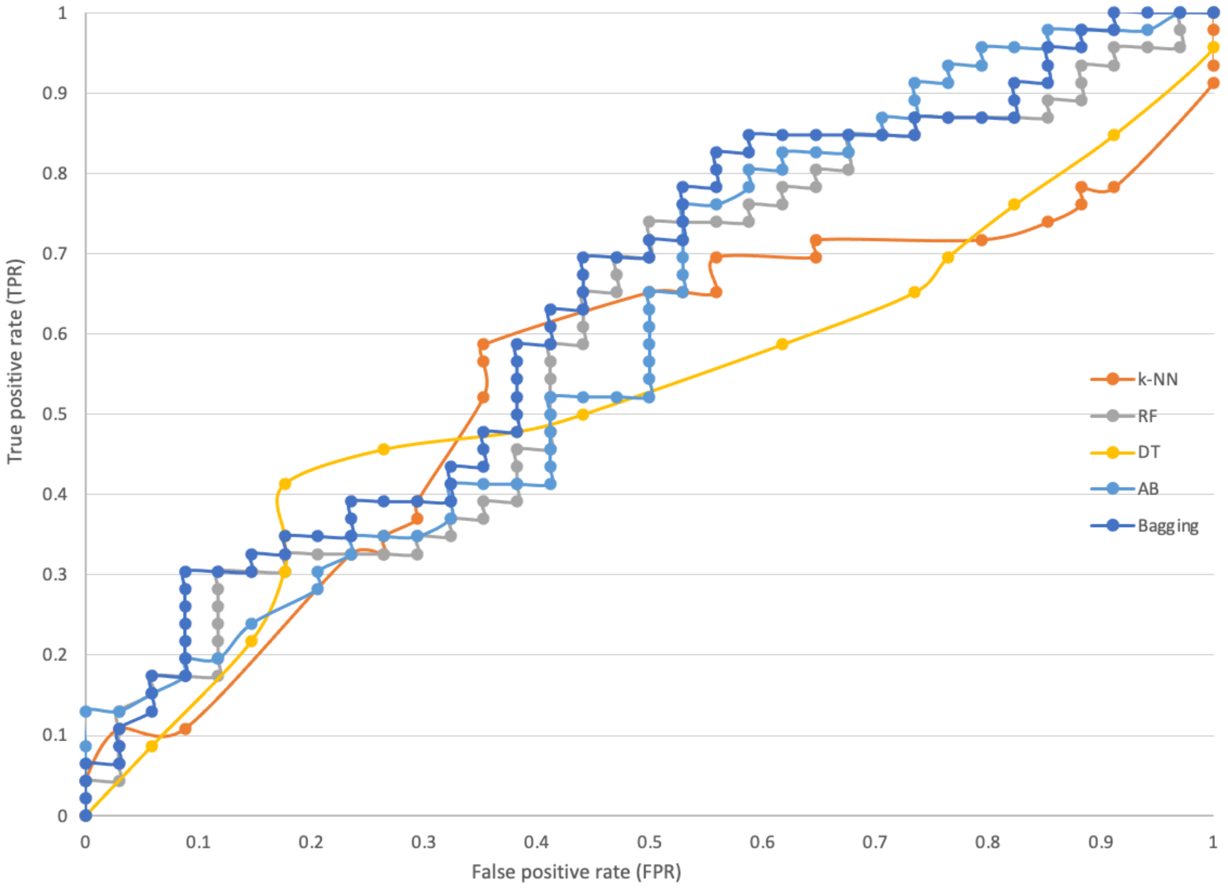
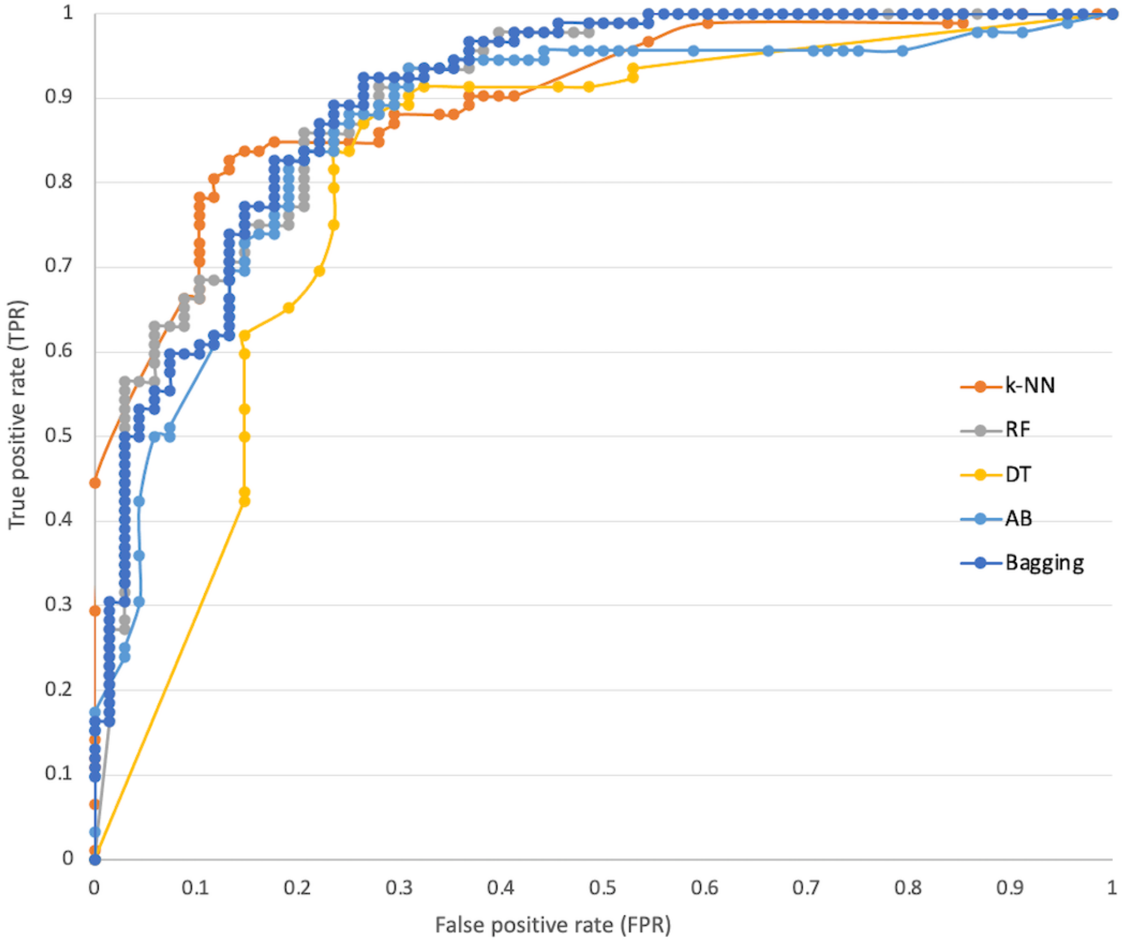


Figure 4. Receiver operating characteristic curves of all classifiers based on enriched data. AB: AdaBoost; DT: decision tree; k-NN: k-nearest neighbor; RF: random forest.

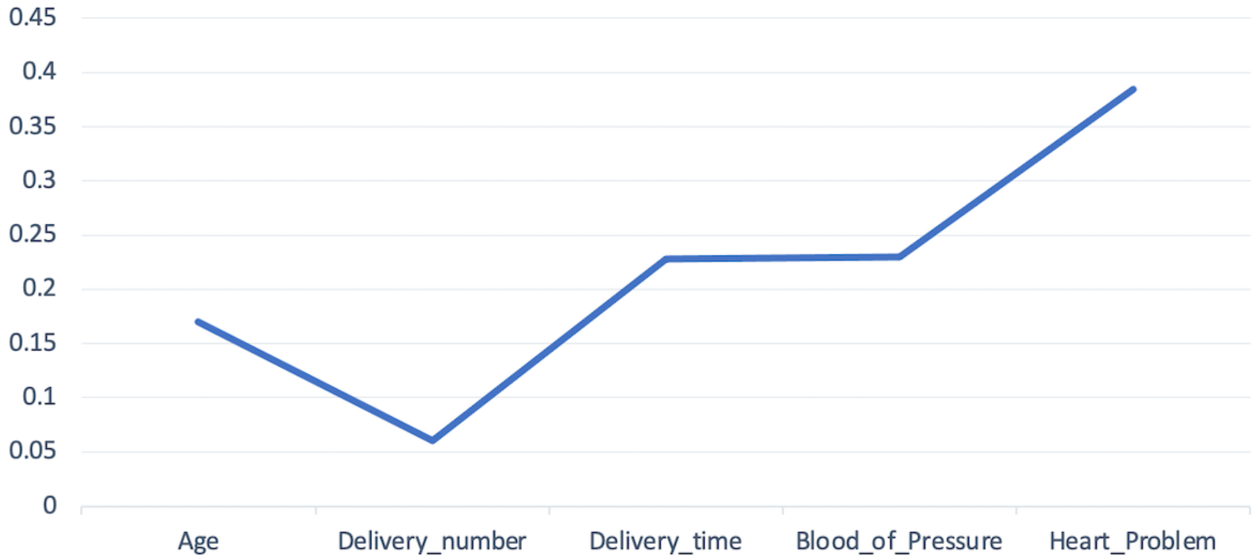


Discussion

Principal Findings

The outcomes in the above tables and figures show that the models' performance in terms of accuracy, sensitivity, specificity, F-measure, and ROC curve is high when trained using the enriched data set compared to the measures achieved using the original data set. In particular, the outcomes shown in Table 1 and Table 3 and Figure 4 represent high model accuracies based on the enriched data set computed using various evaluation methods. All these models were evaluated using cross-validation, which is a commonly adopted method that is considered reliable for models' evaluation in machine learning. In comparison, Amin and Ali [12] and Soleimanian et al [15] trained their models using the same data set (original) and achieved higher accuracy results than those in this study. The reason for achieving higher accuracy results was due to the optimistic method adopted for evaluation using the whole data set as a training set, which was not encouraged in several other studies such as that of Mitchell [48], Smith and Frank [45], and Brownlee [49]. Moreover, this study investigated the relationship of each attribute to its class. A correlation test was performed to identify factors influencing the mode of delivery. In this regard, the relationship of each attribute to its class was estimated. Figure 5 shows the correlation of each attribute to its class. The correlation of each attribute to its class is not high, but on closer investigation, this study concluded that the attribute "heart problem" is strongly correlated with class compared to other attributes, and this factor positively influences the mode of delivery. Thus, a patient with chronic heart-related issues may lead the obstetrician to a decision that is more favorable to c-section than normal delivery.

Figure 5. Correlation between attributes and class.



Conclusion and Future Directions

This study investigated the mode of childbirth by pregnant women by using a machine learning approach. To this end, 5 classification models were trained in order to identify the optimal prediction model to assist obstetricians in decision making for the mode of delivery before birth. In the first part, the original data set was synthesized by populating its records based on the existing ones by using a standard machine learning approach referred to as SMOTE. In the second part, 5 machine learning models were trained based on the original and modified enriched data sets. The models that were trained using the enriched data set performed far better than those trained using the original data set in terms of accuracy, sensitivity, specificity, F-measure, and ROC. This clear difference in the results between the 2 sets of models was due to the increase of records in the original data set. In particular, for the model set trained with the enriched data set, k-NN outperformed the rest of the models with accuracy of 84.38%, while bagging, RF, DT, and AdaBoost showed accuracies of 83.75%, 83.13%, 81.25%, and 80.63%, respectively. Overall, the prediction models developed based on the enriched data set showed similar performances, and therefore the accuracy, sensitivity, specificity, F-measure, and ROC all indicate that these models should be used in the maternity care decision-making process as well as in assisting the obstetrician and midwife in making decisions about the mode of delivery before birth. The data set was artificially populated using a machine learning method. However, in future, if the same data set with the same features enriched with real clinical data will help identify more accurate results, the accuracy may be even more enhanced. The enriched data set in its current stage used in this study yields better results than the original data set, but this data set can become the best if the records are increased with real clinical data.

Acknowledgments

This research work was funded by Institutional Fund Projects under grant no. (IFPHI-272-611-2020). Therefore, the authors gratefully acknowledge technical and financial support from the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

Conflicts of Interest

None declared.

References

1. Sana A, Razzaq S, Ferzund J. Automated Diagnosis and Cause Analysis of Cesarean Section Using Machine Learning Techniques. IJMLC 2012:677-680. [doi: 10.7763/ijmlc.2012.v2.213]
2. Improta G, Ricciardi C, Amato F, D'Addio G, Cesarelli M, Romano M. Efficacy of machine learning in predicting the kind of delivery by cardiotocography. In: IFMBE Proceedings. Cham: Springer; 2019:793-799.
3. Wollmann CL, Hart KD, Liu C, Caughey AB, Stephansson O, Snowden JM. Predicting vaginal birth after previous cesarean: Using machine-learning models and a population-based cohort in Sweden. Acta Obstet Gynecol Scand 2021 Mar;100(3):513-520 [FREE Full text] [doi: 10.1111/aogs.14020] [Medline: 33031579]
4. Fergus P, Selvaraj M, Chalmers C. Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using Cardiotocography traces. Comput Biol Med 2018 Feb 01;93:7-16. [doi: 10.1016/j.compbiomed.2017.12.002] [Medline: 29248699]

5. Pereira S, Portela F, Santos MF, Machado J, Abelha A. Predicting Type of Delivery by Identification of Obstetric Risk Factors through Data Mining. Procedia Computer Science 2015;64:601-609. [doi: 10.1016/j.procs.2015.08.573]
6. Kamat A, Oswal V, Datar M. Implementation of classification algorithms to predict mode of delivery. Int J Comput Sci Inf Technology. 2015. URL: http://ijcsit.com/docs/Volume%206/vol6issue05/ijcsit2015060579.pdf [accessed 2021-05-26]
7. Islam MN, Mahmud T, Khan NI, Mustafina SN, Islam AKMN. Exploring Machine Learning Algorithms to Find the Best Features for Predicting Modes of Childbirth. IEEE Access 2021;9:1680-1692. [doi: 10.1109/access.2020.3045469]
8. Prema NS, Pushpalatha MP. Analysis of association between cesarean delivery and gestational diabetes mellitus using machine learning. Proc Eng Technol Innov. 2020 Apr 27. URL: https://ojs.imeti.org/index.php/PETI/article/view/4740 [accessed 2021-05-26]
9. Abbas SA, Riaz R, Kazmi SZH, Rizvi SS, Kwon SJ. Cause Analysis of Caesarian Sections and Application of Machine Learning Methods for Classification of Birth Data. IEEE Access 2018;6:67555-67561. [doi: 10.1109/access.2018.2879115]
10. Khan N, Mahmud T, Islam M, Mustafina S. Prediction of cesarean childbirth using ensemble machine learning methods. 2020 Presented at: iiWAS '20: Proceedings of the 22nd International Conference on Information Integration and Web-based Applications & Services; November 2020; Thailand p. 331-339. [doi: 10.1145/3428757.3429138]
11. Molina G, Weiser TG, Lipsitz SR, Esquivel MM, Uribe-Leitz T, Azad T, et al. Relationship Between Cesarean Delivery Rate and Maternal and Neonatal Mortality. JAMA 2015 Dec 01;314(21):2263-2270. [doi: 10.1001/jama.2015.15553] [Medline: 26624825]
12. Amin M, Ali A. Performance evaluation of supervised machine learning classifiers for predicting health care operational decisions. C-Section Classification Database Report, UCI Machine Learning Repository. URL: https://api.deepai.org/publication-download-pdf/performance-evaluation-of-supervised-machine-learning-classifiers-for-predicting-healthcare-operational-decisions [accessed 2021-05-26]
13. Reasons for a cesarean birth. American Pregnancy Association. URL: https://americanpregnancy.org/healthy-pregnancy/labor-and-birth/reasons-for-a-cesarean-665/ [accessed 2021-02-16]
14. Li H, Luo S, Trasande L, Hellerstein S, Kang C, Li J, et al. Geographic Variations and Temporal Trends in Cesarean Delivery Rates in China, 2008-2014. JAMA 2017 Jan 03;317(1):69-76. [doi: 10.1001/jama.2016.18663] [Medline: 28030701]
15. Soleimanian F, Mohammadi P, Hakimi P. Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study. IJCA 2012 Aug 30;52(6):21-26. [doi: 10.5120/8206-1613]
16. Sodsee S. Predicting Caesarean Section by Applying Nearest Neighbor Analysis. Procedia Computer Science 2014;31:5-14. [doi: 10.1016/j.procs.2014.05.239]
17. Dulitzki M, Soriano D, Schiff E, Chetrit A, Mashiach S, Seidman DS. Effect of Very Advanced Maternal Age on Pregnancy Outcome and Rate of Cesarean Delivery. Obstetrics & Gynecology 1998;92(6):935-939. [doi: 10.1097/00006250-199812000-00009]
18. Lee KA, Gay CL. Sleep in late pregnancy predicts length of labor and type of delivery. Am J Obstet Gynecol 2004 Dec;191(6):2041-2046. [doi: 10.1016/j.ajog.2004.05.086] [Medline: 15592289]
19. Lipschuetz M, Guedalia J, Rottenstreich A, Novoselsky Persky M, Cohen SM, Kabiri D, et al. Prediction of vaginal birth after cesarean deliveries using machine learning. Am J Obstet Gynecol 2020 Jun;222(6):613.e1-613.e12. [doi: 10.1016/j.ajog.2019.12.267] [Medline: 32007491]
20. Ricciardi C, Improta G, Amato F, Cesarelli G, Romano M. Classifying the type of delivery from cardiotocographic signals: A machine learning approach. Comput Methods Programs Biomed 2020 Nov;196:105712. [doi: 10.1016/j.cmpb.2020.105712] [Medline: 32877811]
21. Saleem S, Naqvi SS, Manzoor T, Saeed A, Ur Rehman N, Mirza J. A Strategy for Classification of "Vaginal vs. Cesarean Section" Delivery: Bivariate Empirical Mode Decomposition of Cardiotocographic Recordings. Front Physiol 2019;10:246 [FREE Full text] [doi: 10.3389/fphys.2019.00246] [Medline: 30941054]
22. Fergus P, Hussain A, Al-Jumeily D, Huang D, Bouguila N. Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms. Biomed Eng Online 2017 Jul 06;16(1):89 [FREE Full text] [doi: 10.1186/s12938-017-0378-z] [Medline: 28679415]
23. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
24. Garner S, Cunningham S, Holmes G, Nevill-Manning C, Witten I. Applying a machine learning workbench: experience with agricultural databases. Research Gate. 1995. URL: https://www.researchgate.net/publication/2827821_Applying_a_Machine_Learning_Workbench_Experience_with_Agricultural_Databases [accessed 2021-05-26]
25. Jamjoom M. The pertinent single-attribute-based classifier for small datasets classification. IJECE 2020 Jun 01;10(3):3227. [doi: 10.11591/ijece.v10i3.pp3227-3234]
26. Jamjoom M, El Hindi K. Partial instance reduction for noise elimination. Pattern Recognition Letters 2016 Apr;74:30-37. [doi: 10.1016/j.patrec.2016.01.021]
27. Caesarian section classification dataset data set. University of California Machine Learning Repository. URL: https://archive.ics.uci.edu/ml/datasets/Caesarian+Section+Classification+Dataset [accessed 2021-03-26]
28. Vapnik V. Statistical Learning Theory. New York: Wiley; 2000.

29. Jiang Z, Pan T, Zhang C, Yang J. A New Oversampling Method Based on the Classification Contribution Degree. Symmetry 2021 Jan 26;13(2):194. [doi: 10.3390/sym13020194]
30. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics 2013 Mar 22;14:106 [FREE Full text] [doi: 10.1186/1471-2105-14-106] [Medline: 23522326]
31. Blagus R, Lusa L. Class prediction for high-dimensional class-imbalanced data. BMC Bioinformatics 2010 Oct 20;11:523 [FREE Full text] [doi: 10.1186/1471-2105-11-523] [Medline: 20961420]
32. Ramezankhani A, Pournik O, Shahrabi J, Azizi F, Hadaegh F, Khalili D. The Impact of Oversampling with SMOTE on the Performance of 3 Classifiers in Prediction of Type 2 Diabetes. Med Decis Making 2014 Dec 01;36(1):137-144. [doi: 10.1177/0272989x14560647]
33. Alghamdi M, Al-Mallah M, Keteyian S, Brawner C, Ehrman J, Sakr S. Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) project. PLoS One 2017;12(7):e0179805 [FREE Full text] [doi: 10.1371/journal.pone.0179805] [Medline: 28738059]
34. Mohammed AJ, Hassan MM, Kadir DH. Improving Classification Performance for a Novel Imbalanced Medical Dataset using SMOTE Method. IJATCSE 2020 Jun 25;9(3):3161-3172. [doi: 10.30534/ijatcse/2020/104932020]
35. Shuja M, Mittal S, Zaman M. Effective Prediction of Type II Diabetes Mellitus Using Data Mining Classifiers and SMOTE. Adv Comput Intell Syst Algorithms Intell Syst 2020.? 2020 Jan 03:195-211. [doi: 10.1007/978-981-15-0222-4_17]
36. Nnamoko N, Korkontzelos I. Efficient treatment of outliers and class imbalance for diabetes prediction. Artif Intell Med 2020 Apr;104:101815 [FREE Full text] [doi: 10.1016/j.artmed.2020.101815] [Medline: 32498997]
37. Hussain L, Lone KJ, Awan IA, Abbasi AA, Pirzada J. Detecting congestive heart failure by extracting multimodal features with synthetic minority oversampling technique (SMOTE) for imbalanced data using robust machine learning techniques. Waves in Random and Complex Media 2020 Aug 24:1-24. [doi: 10.1080/17455030.2020.1810364]
38. Ebenuwa S. Re: Which are the best methods to artificially increase the size of numeric data sets? Research Gate. URL: https://www.researchgate.net/post/Which-are-the-best-methods-to-artificially-increase-the-size-of-numeric-datasets [accessed 2021-03-02]
39. Frank E. Re: Data augmentation in deep learning. Weka. URL: https://weka.8497.n7.nabble.com/Data-augmentation-in-deep-learning-td42467.html [accessed 2021-03-02]
40. Song Y, Lu Y. Decision tree methods: applications for classification and prediction. Shanghai Arch Psychiatry 2015;27(2):130-135. [doi: 10.11919/j.issn.1002-0829.215044] [Medline: 26120265]
41. Breiman L. Random forests. Mach Learn 2001;45:5-32. [doi: 10.1023/A:1010933404324]
42. Zhang C. Ensemble Machine Learning: Methods and Applications. New York: Springer-Verlag; 2012.
43. Datta S, Pihur V, Datta S. An adaptive optimal ensemble classifier via bagging and rank aggregation with applications to high dimensional data. BMC Bioinformatics 2010 Aug 18;11:427 [FREE Full text] [doi: 10.1186/1471-2105-11-427] [Medline: 20716381]
44. Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. Mach Learn 1991 Jan;6(1):37-66. [doi: 10.1007/bf00153759]
45. Smith T, Frank E. Introducing machine learning concepts with WEKA. In: Statistical Genomics. Methods in Molecular Biology. New York: Humana Press; 2016:353-378.
46. Geisser S. The Predictive Sample Reuse Method with Applications. Journal of the American Statistical Association 1975;70(350):320-328. [doi: 10.1080/01621459.1975.10479865]
47. Attwal K, Dhiman A. Exploring data mining tool - Weka and using Weka to build and evaluate predictive models. Adv Appl Math Sci. 2020. URL: https://www.mililink.com/upload/article/131953194aams_vol_196_april_2020_a3_p451-469_kanwal_preet_singh_attwal.pdf [accessed 2021-05-27]
48. Mitchell T. Machine Learning. NY, USA: McGraw Hill; 1997.
49. Brownlee J. Machine Learning Mastery With Weka. URL: http://scholar.googleusercontent.com/scholar?q=cache:dsK4nil5JkUJ:scholar.google.com/+Machine+Learning+Mastery+With+Weka&hl=en&as_sdt=0,5 [accessed 2021-03-02]
50. Wong T. Parametric methods for comparing the performance of two classification algorithms evaluated by k-fold cross validation on multiple data sets. Pattern Recognition 2017 May;65:97-107. [doi: 10.1016/j.patcog.2016.12.018]
51. Anguita D, Ghelardoni L, Ghio A, Oneto L, Ridella S. The ‘K’ in k-fold cross validation. In: Comput Intell Mach Learn. 2012 Presented at: 20th Eur Symp Artif Neural Networks; 2012; Bruges (Belgium) p. 441-446.
52. Schaffer C. Selecting a classification method by cross-validation. Mach Learn 1993 Oct;13(1):135-143. [doi: 10.1007/bf00993106]
53. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
54. McHugh M. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
55. Parson D. Kappa statistic. Data Mining and Predictive Analytics. URL: http://faculty.kutztown.edu/parson/fall2019/Fall2019Kappa.html#:~:text=?The [accessed 2021-02-15]

56. Powers DMW. Evaluation: from precision, recall F-measure to ROC informedness, markedness and correlation. International Journal of Machine Learning Technology. 2011. URL: http://arxiv.org/abs/2010.16061 [accessed 2021-03-02]
57. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care & Pain 2008 Dec;8(6):221-223. [doi: 10.1093/bjaceaccp/mkn041]
58. Van Rijsbergen C. Information Retrieval. 2nd edition. London: Butterworth; 1979.
59. Sasaki Y. The Truth of the F-measure. 2007. URL: https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf [accessed 2021-05-26]
60. Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. Invest Radiol 1979;14(2):109-121. [doi: 10.1097/00004424-197903000-00002] [Medline: 478799]
61. Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006 Jun;27(8):861-874. [doi: 10.1016/j.patrec.2005.10.010]
62. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr 2011 Apr;48(4):277-287. [doi: 10.1007/s13312-011-0055-4] [Medline: 21532099]
63. Vuk M, Curk T. ROC curve, lift chart and calibration plot. Metodoloski zvezki. 2006. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.126.7382&rep=rep1&type=pdf [accessed 2021-05-26]

Abbreviations

C-section: cesarean section
DT: decision tree
k-NN: k-nearest neighbor
RF: random forest
ROC: receiver operating characteristic
SMOTE: Synthetic Minority Oversampling Technique

Edited by R Kukafka; submitted 16.03.21; peer-reviewed by A Alrasheed, K Fakeeh; comments to author 28.03.21; revised version received 30.03.21; accepted 30.04.21; published 04.06.21.

Please cite as:
Ullah Z, Saleem F, Jamjoom M, Fakieh B
Reliable Prediction Models Based on Enriched Data for Identifying the Mode of Childbirth by Using Machine Learning Methods: Development Study
J Med Internet Res 2021;23(6):e28856
URL: https://www.jmir.org/2021/6/e28856
doi:10.2196/28856
PMID:34085938

Original Paper

Reduction of Time on the Ground Related to Real-Time Video Detection of Falls in Memory Care Facilities: Observational Study

Eleonore Bayen[1,2], MD, PhD; Shirley Nickels[3], BSc; Glen Xiong[4], MD; Julien Jacquemot[3], BSc, MSc; Raghav Subramaniam[3], BSc, MSc; Pulkit Agrawal[3,5], PhD; Raheema Hemraj[3], BA, BSc, MSc; Alexandre Bayen[6], PhD; Bruce L Miller[2], MD; George Netscher[3], BSc, MSc

[1]Department of Neuro-rehabilitation, Hôpital Pitié-Salpêtrière, Assistance Publique des Hôpitaux de Paris, Sorbonne Université, Paris, France
[2]Global Brain Health Institute, Memory and Aging Center, Department of Neurology, University of California, San Francisco, CA, United States
[3]SafelyYou, Inc, San Francisco, CA, United States
[4]Alzheimer's Disease Center, Department of Neurology, University of California, Davis, Sacramento, CA, United States
[5]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Boston, CA, United States
[6]Electrical Engineering and Computer Science Department, University of California, Berkeley, CA, United States

Corresponding Author:
Eleonore Bayen, MD, PhD
Department of Neuro-rehabilitation
Hôpital Pitié-Salpêtrière
Assistance Publique des Hôpitaux de Paris, Sorbonne Université
47 Bd de l'Hôpital
Paris, 75013
France
Phone: 33 142161101
Fax: 33 142161148
Email: eleonore.bayen@gbhi.org

Abstract

Background: Lying on the floor for a long period of time has been described as a critical determinant of prognosis following a fall. In addition to fall-related injuries due to the trauma itself, prolonged immobilization on the floor results in a wide range of comorbidities and may double the risk of death in elderly. Thus, reducing the length of Time On the Ground (TOG) in fallers seems crucial in vulnerable individuals with cognitive disorders who cannot get up independently.

Objective: This study aimed to examine the effect of a new technology called SafelyYou Guardian (SYG) on early post-fall care including reduction of Time Until staff Assistance (TUA) and TOG.

Methods: SYG uses continuous video monitoring, artificial intelligence, secure networks, and customized computer applications to detect and notify caregivers about falls in real time while providing immediate access to video footage of falls. The present observational study was conducted in 6 California memory care facilities where SYG was installed in bedrooms of consenting residents and families. Fall events were video recorded over 10 months. During the baseline installation period (November 2017 to December 2017), SYG video captures of falls were not provided on a regular basis to facility staff review. During a second period (January 2018 to April 2018), video captures were delivered to facility staff on a regular weekly basis. During the third period (May 2018 to August 2018), real-time notification (RTN) of any fall was provided to facility staff. Two digital markers (TUA, TOG) were automatically measured and compared between the baseline period (first 2 months) and the RTN period (last 4 months). The total number of falls including those happening outside of the bedroom (such as common areas and bathrooms) was separately reported by facility staff.

Results: A total of 436 falls were recorded in 66 participants suffering from Alzheimer disease or related dementias (mean age 87 years; minimum 65, maximum 104 years). Over 80% of the falls happened in bedrooms, with two-thirds occurring overnight (8 PM to 8 AM). While only 8.1% (22/272) of falls were scored as moderate or severe, fallers were not able to stand up alone in 97.6% (247/253) of the cases. Reductions of 28.3 (CI 19.6-37.1) minutes in TUA and 29.6 (CI 20.3-38.9) minutes in TOG were observed between the baseline and RTN periods. The proportion of fallers with TOG >1 hour fell from 31% (8/26; baseline) to zero events (RTN period). During the RTN period, 76.6% (108/141) of fallers received human staff assistance in less than 10 minutes, and 55.3% (78/141) of them spent less than 10 minutes on the ground.

Conclusions: SYG technology is capable of reducing TOG and TUA while efficiently covering the area (bedroom) and time zone (nighttime) that are at highest risk. After 6 months of SYG monitoring, TOG was reduced by a factor of 3. The drastic reduction of TOG is likely to decrease secondary comorbid complications, improve post-fall prognosis, and reduce health care costs.

(J Med Internet Res 2021;23(6):e17551) doi:10.2196/17551

KEYWORDS

artificial intelligence; video monitoring; real-time video detection; fall; time on the ground; Alzheimer disease; dementia; memory care facilities

Introduction

Falls are the leading cause of injuries among people aged 65 and older, with estimated yearly direct medical costs of US $637.2 million for fatal falls and US $31.3 billion for nonfatal falls in the United States [1]. Older adults with cognitive impairment have an increased risk of falling [2], and those with dementia living in nursing care facilities fall 4.1 times per year on average versus 2.3 times for other residents [3]. Additionally, individuals with dementia are the least likely to call for assistance when they cannot rise alone from the ground, and cognitive impairment is found to be the most significant factor that predicts lying on the floor for a long time after a fall [4].

Reducing the length of time on the floor after a fall is crucial because fall-related injuries include not only traumatic injuries associated with the acute fall but also comorbid complications related to prolonged post-fall immobilization. Artificial intelligence (AI) in the field of computer vision; innovative health technology using secure, network-attached storage; and a customized computer application together offer the potential to detect falls rapidly without the need for wearable devices and to support care in dementia care facilities [5]. In previous work using SafelyYou Guardian (SYG) technology [5], we showed that video monitoring of falls along with regular fall review by facility staff resulted in more accurate identification of falls, fall mechanisms, and the injuries related to the trauma itself (such as traumatic brain injury, for instance) [6]. More recently, we compared nonbeneficiaries to beneficiaries of SYG and found that review of fall footage enabled a significant reduction of visits to the emergency department, by 72% on average, due to better triaging of fall-related injuries [7]. In this article, we focus on subsequent length of time spent on the floor after falling as automatically measured by SYG technology in beneficiaries of SYG.

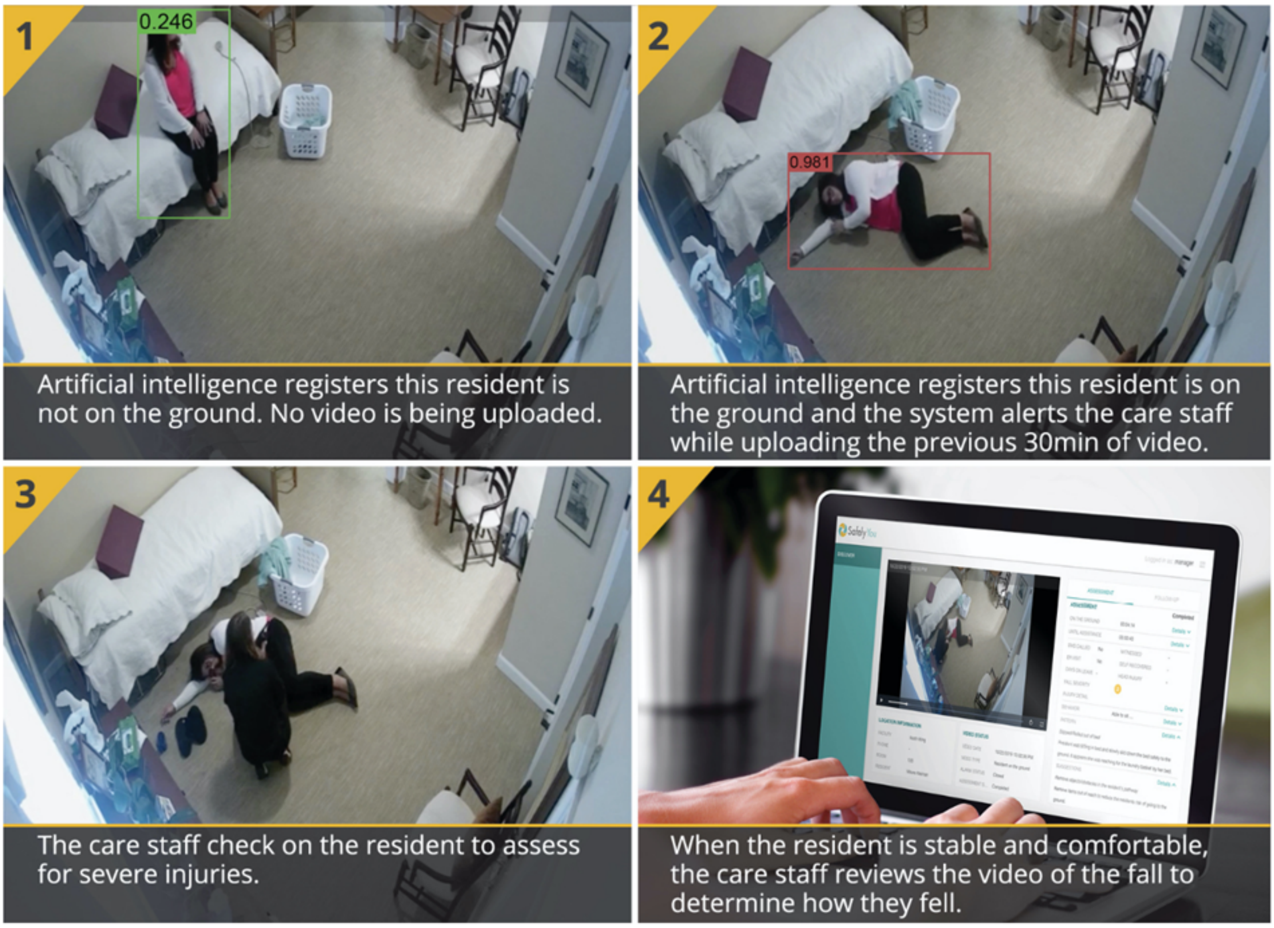
The objective here was to investigate the effect of the SYG real-time notification (RTN) system on early provision of post-fall care. We analyzed 2 lengths of time associated with a fall event: (1) The length of Time Until staff Assistance arrives (TUA) reflects the first support provided by professional caregivers, and (2) the length of Time On the Ground (TOG) in fallers is a prognostic marker reflecting risks for secondary post-fall complications in residents. Discrepancies between TUA and TOG are common. For example, a professional caregiver who arrives in a resident's room shortly after a fall may wait for a second caregiver to come and assist the resident to his or her bed. These 2 digital care markers were measured over 10 months for consenting residents living in 6 memory care facilities. We hypothesized that introducing SYG would reduce TUA and TOG. Comparison of TUA and TOG markers between the baseline period and RTN period were computed to assess the effect of SYG technology on early post-fall care.

Methods

Design of the Study and Technology Equipment

This study is part of a larger project involving multilevel collaborations between engineers, researchers, clinicians, and care providers. SafelyYou develops computer vision algorithms (a subfield of AI focused on visual understanding) for automated, real-life, real-time detection and notification of falls in memory care facilities [5]. This observational study was carried out during 10 months in 6 memory care facilities in California, and a progressive introduction of SYG technology was carried out so that facility staff could get familiar with the technology. In each of the 6 facilities, wall-mounted cameras were installed in residents' bedrooms of consenting elderly and families (at the exclusion of residents' personal bathrooms, where no camera was placed) and in accordance with the privacy and ethical guidelines discussed below. There was no requirement about where the camera was placed, but cameras were generally placed in order to capture as much of the room as possible since the camera would not detect a fall that it could not see. The field of view of the camera was greater than 90 degrees such that an entire room could be captured if the camera was placed in an upper corner of the room. A fall was defined as an "unexpected event in which the participant comes to rest on the ground, floor, or lower level" [8]. A fall incident was detected by the AI-enabled camera system when a resident was identified on the floor, whatever his or her position (sitting or lying). This detection is independent of camera position, the number of individuals in the room, and which individual is on the floor. "On the floor" was defined to be when one body part other than the feet touched the floor, such as a knee, forearm, or the posterior. Falls were video captured 24 hours a day, 7 days a week, and video data were securely transmitted using a local network storage device. A secure, customized computer application gave facility staff access to the video footage and fall information of fall incidents (Figure 1).

Figure 1. SafelyYou Guardian (SYG) technology provides automated notification of the fall, video footage (to assess fall severity), and care recommendations to facility staff.



Artificial intelligence registers this resident is not on the ground. No video is being uploaded.

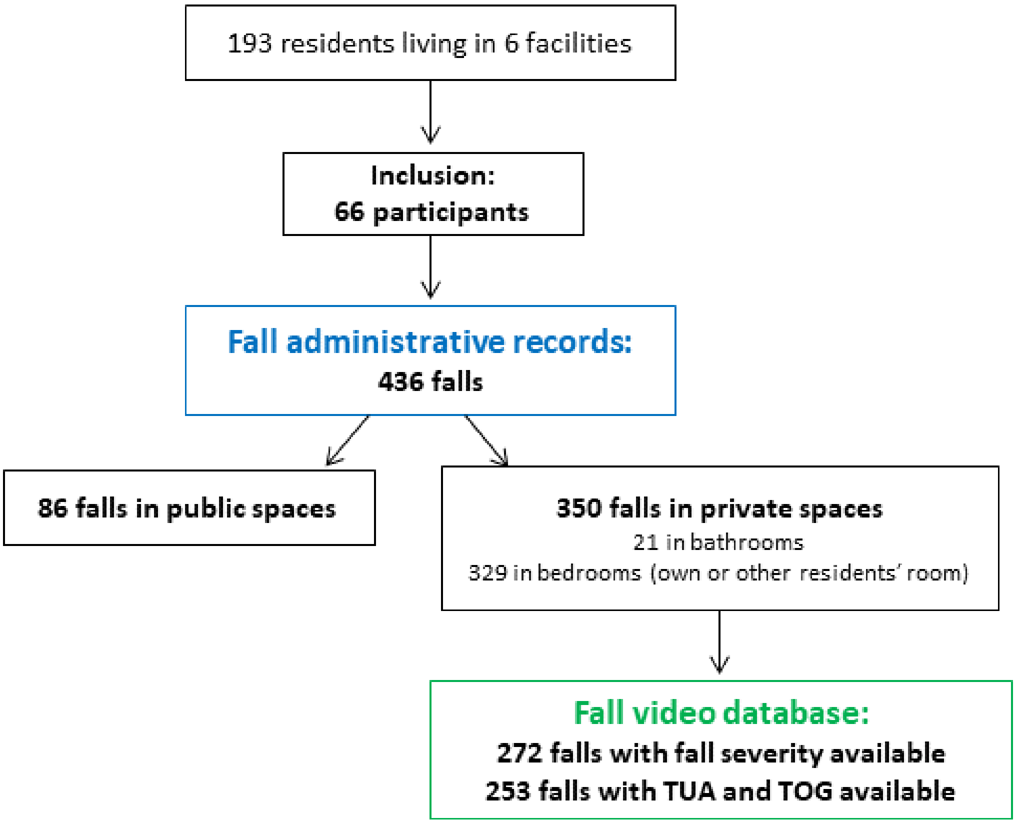
Artificial intelligence registers this resident is on the ground and the system alerts the care staff while uploading the previous 30min of video.

The care staff check on the resident to assess for severe injuries.

When the resident is stable and comfortable, the care staff reviews the video of the fall to determine how they fell.

Between November 1, 2017 and August 31, 2018, falls were prospectively recorded in administrative reports and were video captured for consenting participants (Figure 2). The study entailed 3 successive periods over 10 months. During the 2-month baseline installation period (November 2017 and December 2017), a set of control measures was performed following setup of the cameras and SYG video captured falls, but video footage was not provided on a regular basis to facility staff review. During a second 4-month period (January 2018 through April 2018), video captures of falls were delivered to facility staff on a weekly basis so that they could get familiar with the technology and footage review; however, RTN directly occurring after a fall was not activated. Finally, during the third 4-month period (from May 2018 through August 2018), RTN of any fall was provided to facility staff via a phone call so that the faller could receive immediate assistance after the alarm onset.

Figure 2. Flow chart showing inclusions and fall databases. TOG: Time On the Ground; TUA: Time Until staff Assistance arrives.



Outcome Measures

For each participant, the number of falls occurring in private and public spaces of the memory care facility was recorded by facility staff over the 10-month period: Private spaces refer to resident's personal areas (ie, individual's bedroom and bathroom), while public spaces refer to common areas shared by residents (eg, dining rooms, hallways, lounges). Falls happening in bathrooms, in nonparticipants' bedrooms, and in common spaces of the memory care facilities were not video recorded but were reported through existing incident reporting processes.

For participants only, a video review allowed the classification of falls according to their severity. A fall was classified as a "behavioral fall" when the resident showed an intentional slow descent with safe landing but no recovery to prior position. Other falls were granted a severity score using the 4-point Hopkins Falls Grading Scale [9]. This scale stratifies fall severity in near fall (Grade 1), minor fall with no need for medical attention (Grade 2), moderate fall requiring medical attention (Grade 3), and severe fall requiring hospital admission (Grade 4).

Two care markers were automatically generated by the SYG technology in order to assess early provision of post-fall care. TUA timed the delay between the detection of the fall and the detection of staff arrival in the resident's bedroom. TOG timed the delay between the detection of the fall and the resident's return to a comfortable position with assistance from facility staff. Measurement of these 2 digital markers allowed classification of falls according to their duration [4] (less than 10 minutes, 10 to 30 minutes, 30 to 60 minutes, and more than 60 minutes).

Ethical Procedures and Privacy

Privacy and consent procedures were developed with support from the Institutional Review Board of the University of California, Berkeley [10] and in collaboration with the California Department of Social Services Community Care Licensing Division. The Committee for Protection of Human Subjects of the University of California, Berkeley approved the study protocol. Procedures have been published in detail [6].

Residents living within the care facilities showed severe cognitive impairment related to Alzheimer disease and related dementias. As a consequence, surrogate consent by a legally authorized representative (usually next of kin) was required for the present study. The legally authorized representatives of the residents were given oral and written detailed information about the study and were provided the informed consent document. The study was explained to the affected individuals living in the facility, and if they provided any verbal or nonverbal indication that they did not wish to have the cameras in their bedroom, they were not included. The legally authorized representative was the person who could say yes to the study, thus providing informed consent, but the resident retained the right to decline the study at any time, thus providing assent. Cameras were located high in a corner in the bedroom, but not in the bathroom, and remained visible to the participants. Camera audio recording was disabled. Signs were posted visibly on the door of each bedroom in which video recording occurred as a reminder to residents, families, and facility staff that participants were being video monitored in the bedrooms.

Statistical Analyses

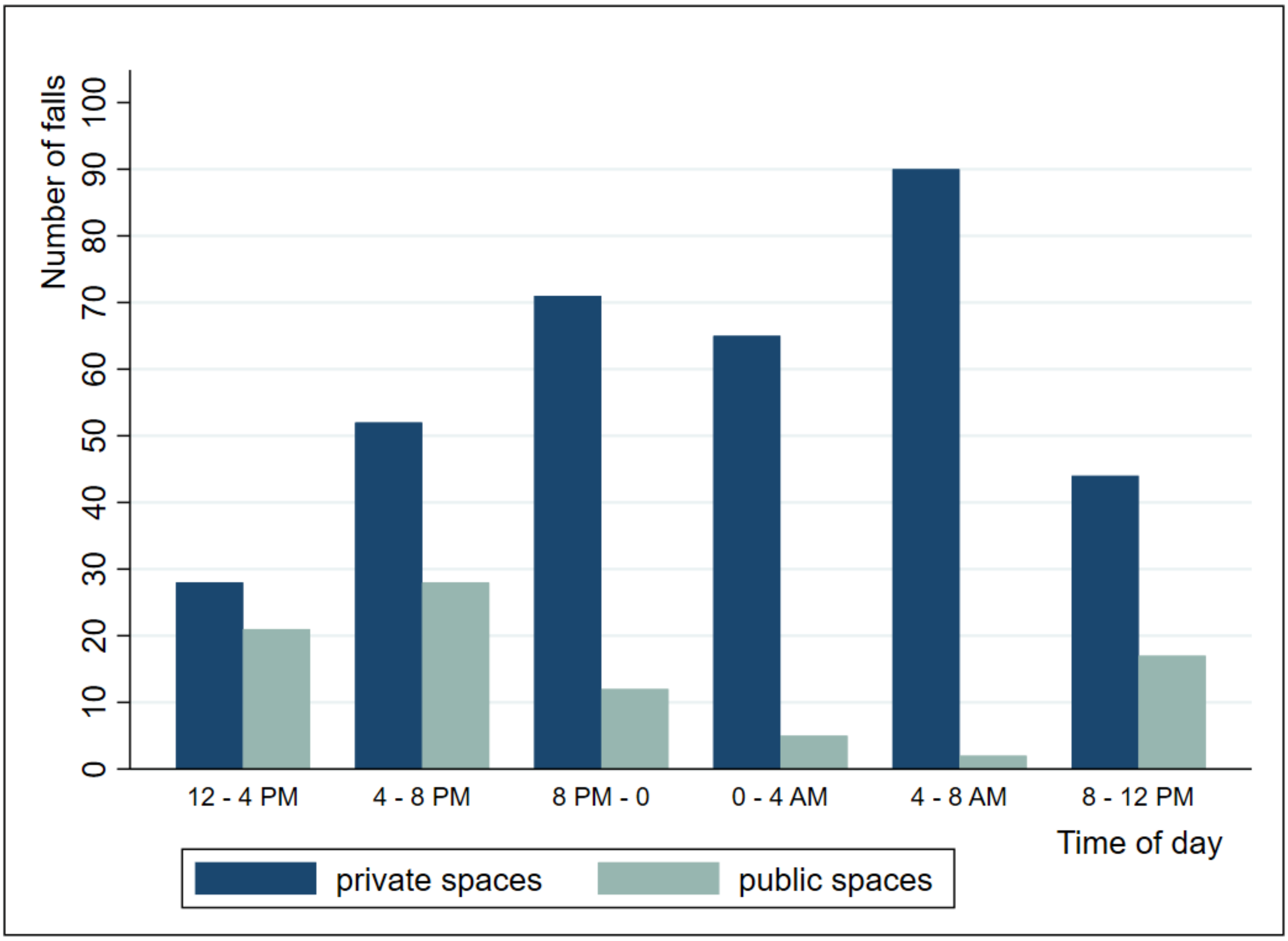
TUA and TOG variables were quantified in minutes and are described with means, median, standard deviation, minimum, and maximum. Comparison of means of TUA and of TOG between the baseline period and the RTN period was performed with nonparametric Wilcoxon rank-sum tests because of nonnormality of the data. A one-way analysis of variance (ANOVA) was conducted to determine if TUA differed between facilities. The TUA amounted to 0 minutes when a facility staff was already in the bedroom witnessing the live fall. Falls with a TUA equal to 0 were counted over the 3 periods, and Wilcoxon rank-sum comparison tests were performed with and without falls with TUA equal to 0. Statistical analyses were performed using Stata version 15 (Stata Corp; College Station, Texas) and R (R Programming).

Results

A total of 66 individuals out of 193 residents (34.2%) living in the 6 facilities at the time of inclusion participated in the study. These 66 participants suffered from Alzheimer disease and related dementias. They had a mean age of 87 (SD 6.9; minimum 65, maximum 104) years and were nearly two-thirds women (42/66, 64%). The majority of the participants (54/66, 82%) were determined to be recurrent fallers with an average of 6.6 (SD 8.6; minimum 1, maximum 49) falls per individual over the 10-month period. No difference in the number of falls per faller was found according to gender.

A total of 436 falls were reported in administrative records of those 66 study participants over 10 months. Among these falls, 86 (86/436, 19.7%) occurred in common areas, and 350 (350/436, 80.3%) occurred in private spaces (including falls in participants' personal bedrooms, in participants' personal bathrooms, and in bedrooms of other nonparticipant residents). As displayed in Figure 3, 64.6% (226/350) of falls that happened in private spaces occurred during nighttime (8 pm to 8 am), with a maximal spike of incidents between 4 am and 8 am. Video data about fall severity were fully available for 272 events, including a total of 93 falls (93/272, 34.2%) classified as "behavioral falls with safe landing." Grading of the severity of the remaining falls showed 18 (18/272, 6.6%) near falls with independent recovery, 139 (139/272, 51.1%) minor falls, and 22 (22/272, 8.1%) moderate and severe falls (18 and 4, respectively).

Figure 3. Display of falls according to location and day schedule. Private spaces refer to residents' personal areas (ie, individual's bedroom and bathroom); public spaces refer to common areas shared by residents (eg, dining rooms, hallways, lounges).

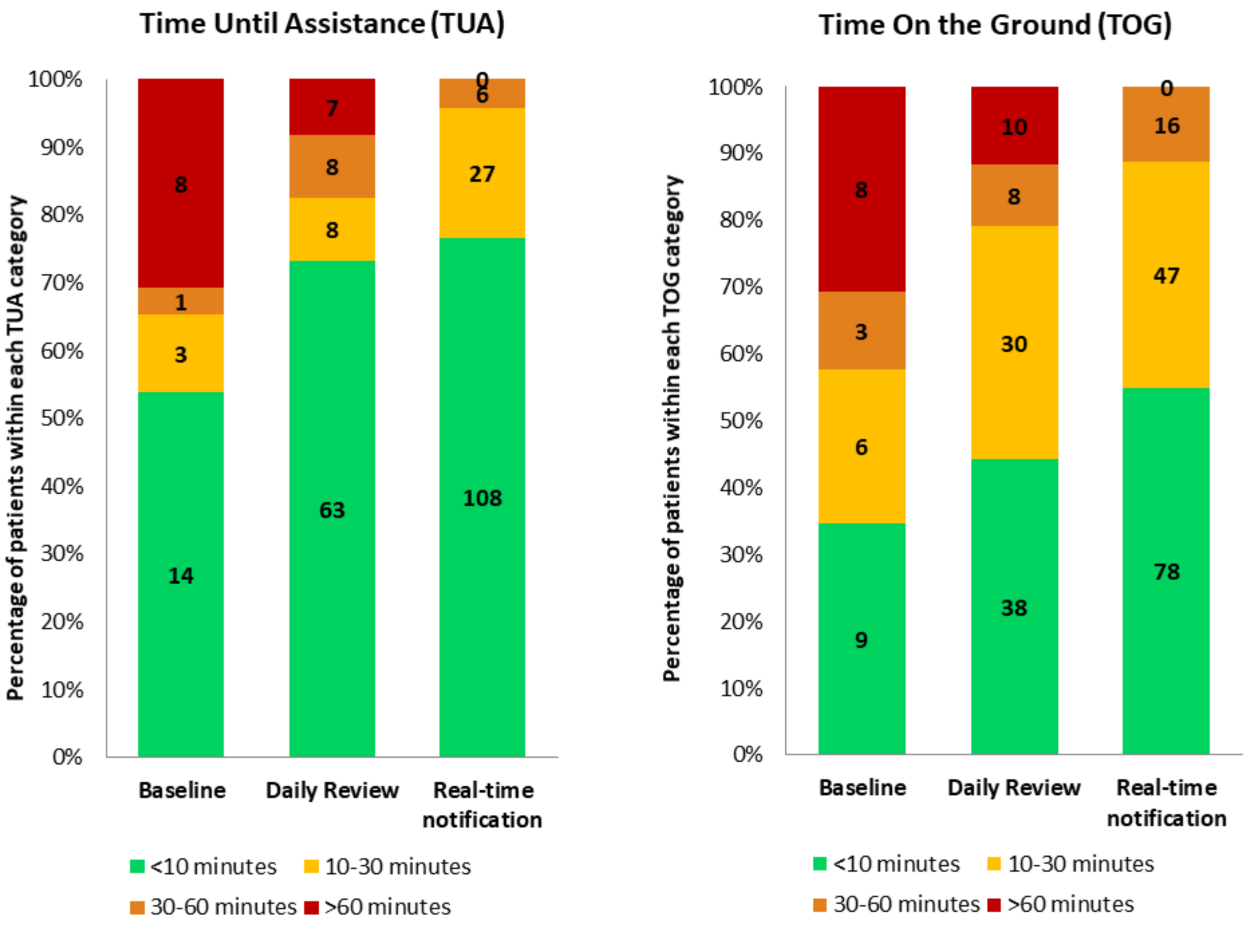


Video data about TUA and TOG markers were fully available in 253 falls (missing videos and gaps in video are attributed to technical issues such as gaps in WiFi connectivity). Both TUA and TOG showed a wide variation ranging from zero (or a few seconds) to more than 3.5 hours (maximum TUA: 214.2 minutes; maximum TOG: 215 minutes; both during the 2 first periods before RTN). The one-way ANOVA showed that there was a statistically significant difference in TUA between facilities (F_5=3.79, *P*=.003). We found that TUA was higher in the standalone memory care facilities (2 of the 6 facilities) than in the other 4 facilities (ie, memory care units within larger assisted-living facilities; Wilcoxon rank-sum test *P*=.002). The difference between TOG and TUA over 10 months was similar over the 3 time periods (7.8 minutes at baseline, 6.6 minutes during weekly review, and 6.9 minutes during the RTN period) and often related to the need for more than one health care professional to help lift fallers. In 2.4% (6/253) of falls, residents got up independently before staff arrival, resulting in a larger TUA than TOG. A facility staff member was already in the bedroom during the fall event for 17.4% (44/253) of cases, resulting in a TUA of 0 minutes.

The TUA amounted to an average 35.3 (SD 48.3, median 7.6) minutes, 15.9 (SD 32.1, median 2.4) minutes, and 7.0 (SD 9.5, median 2.8) minutes during the baseline, weekly review, and RTN periods, respectively. TUA differed significantly between the baseline period (November 2017 and December 2017) and the RTN period (May 2018 to August 2018; *P*=.004), with a mean reduction in TUA of 28.3 (CI 19.6-37.1) minutes. The TOG amounted to an average 43.2 (SD 48.8, median 23.3) minutes, 22.9 (SD 32.9, median 11.5) minutes, and 13.6 (SD 12.1, median 9.3) minutes during the baseline, weekly review, and RTN periods, respectively. TOG differed significantly between the baseline period and RTN period (*P*=.043), with a mean reduction in TOG of 29.6 (CI 20.3-38.9) minutes. We found similar results after removing those falls with a TUA equal to 0 minutes (ie, a significant difference in TUA between the baseline period and RTN period [*P*=.004], with a mean reduction in TUA of 37.7 [CI 27.8-47.7] minutes).

As shown in Figure 4, there was an increase in the proportion of individuals lying on the floor for less than 10 minutes and a related decrease in the proportion of individuals experiencing a TOG of over 60 minutes, with no such case during the RTN period. During the RTN period, 76.6% (108/141) of fallers received human assistance in less than 10 minutes and 55.3% (78/141) of them spent less than 10 minutes on the ground. In parallel, there was a drastic drop from 31% (8/26) to zero cases of those individuals who experienced a TOG of more than 1 hour (Figure 4).

Figure 4. Proportion of falls with short, medium, and long (A) Time Until staff Assistance arrives (TUA) and (B) Time On the Ground (TOG) measures during the 3 time periods.



Discussion

The present observational study provides novel insight about real-life falls and early post-fall care as captured through continuous video monitoring, AI detection, and notification of falls in bedrooms in memory care facility settings. We found that a significant reduction (0.5 hour) in TUA and TOG was associated with video monitoring and RTN of falls. During RTN, three-quarters of fallers received staff assistance in less than 10 minutes, and half of them recovered to a comfortable position in less than 10 minutes, while the proportion of very long TOG (ie, more than 60 minutes) was reduced. These results suggest that SYG technology was able to reduce a crucial severity marker of falls (ie, prolonged immobilization on the ground) that is usually associated with medical complications and poor post-fall prognosis.

The study provides ecological data of falls happening in memory care facilities using both administrative records and video monitoring reports. Regarding locations of falls, over 80% of the fall incidents happened in private spaces (mostly residents' bedrooms) confirming a similar finding that 75% of falls happen in bedrooms, as estimated from 70,000 fall incidents recorded administratively in 528 long-term care facilities [11]. Since residents are usually alone when they are in their bedrooms (as opposed to common areas where more facility staff can assist them), it appears necessary to place fall detector systems like SYG primarily in these high-risk areas to ensure continuous fall monitoring [4]. Regarding time of the day when residents fall, we found a high occurrence during nighttime (when there are fewer facility staff and residents spend more time on their own in their rooms), suggesting potential for adapting schedules and safety rounds in order to minimize risk of falls. Indeed, sleep disturbances (such as reduced sleep time, sleep fragmentation, and nocturnal wandering [12]), impaired circadian rhythmicity with sundowning [13], nocturnal effects of drugs [14], and urination disorders [15] are common in dementia and yield an augmented risk of night falls. Regarding the severity of falls, our results parallel previous research reporting that the vast majority of falls are classified as minor falls and do not include traumatic injuries [16,17]. In turn, this result points out the benefit of reviewing video footage in order to better screen injuries and avoid unnecessary referral to the emergency department and unnecessary health care costs [7]. The review also allows identification of the mechanisms of falls [18]. In long-term care facilities, most falls were related to self-induced weight shifting (with an equal frequency during walking, transferring, and standing) that can further lead to more balance assessment and fall prevention [18].

The study provides quantitative post-fall measures (TUA, TOG) that were automatically measured with an AI-driven technology. To our knowledge, our group is one of the first to report these automatically derived time measures. By contrast, research in the field relies primarily on surveys and retrospective interviews of patients and proxies, leading to recall bias [19] and estimation errors [4,13,16,20]. Our baseline result of 30.8% of residents lying on the floor for more than 1 hour (TOG) replicates a similar proportion of 30% stemming from a large survey of 265 falls from a mixed population of home-dwelling and institutionalized elderly people over 90 years old [4]. For the first time, we were able to show that this major and frequent risk of long TOG events could be successfully modified in just a few months and was reduced down to zero events. We found a TUA logically lower than TOG in more than 93% of the cases; as previously reported, 80% of vulnerable individuals are unable to get up without help, and getting up from the floor often requires more than one health professional [4]. The same study also showed that 94% of institutionalized elderly who could not get up alone and who had access to a call alarm did not activate the alarm system to summon help [4]. In geriatric patients visiting the emergency department after a fall, only 16% of fallers used their personal emergency response system to call for assistance [20]. Our results highlight how AI fall detector systems like SYG do not require a faller's direct action to activate staff assistance [20], which is necessary to deliver a critical care need and enhance safety for vulnerable individuals. Variations in TUA might be related to different facility protocols and staff availability, and caregivers also show diverse experience and training. In turn, facility staff began to incorporate the video review in their care practice after an adoption period (accounting for a reduction of TUA during the second period), and policy changes might also be observed (such as additional safety rounds for residents at high risk and environmental changes as previously documented [6]).

TUA and TOG care markers showed an average 29-minute decrease, indicating that much faster assistance to help fallers recover to a comfortable position could be provided, instead of relying on regular care rounds only. There is extensive literature about the additional critical comorbidities that are related to the long lie [21] and inability to get up [22], including metabolic and physical consequences (dehydration, hypothermia, rhabdomyolysis, renal failure, pressure ulcers [23]), psychological complications (post-fall syndrome with fear of falling again [24] and activity limitation [16]), higher risk for recurrent falls [4], and loss of autonomy [16]. Lying on the floor after a fall for a long period of time was also found to nearly double the risk of death in elderly adults [23]. On average, the RTN resulted here in assigning a "fall incident" to staff as a red flag priority among their other ongoing care tasks, accounting for the substantial reduction in TUA and in TOG.

TOG has been long considered a core severity factor of falls as well as a predictor of post-fall prognosis [22], while it has suffered from quantification weaknesses at the same time [4]. We demonstrate here that AI technology can overcome this issue and can now generate a robust and reliable care marker. In turn, shorter TUA could be considered a quality-of-care criterion for patients, families, and care providers, while TOG could be integrated more systematically in falls reporting. Taking into account that fall guidelines and management entail standardization needs [8,25-27], our contribution to the field is to provide evidence for the feasibility and usefulness to integrate these digital care markers into dementia care practice. Other strategies in the field of fall detection use different technology including wearable alert systems [28] (which require residents and staff to remember to wear and reload the device regularly), accelerometer-based fall detection [29], and nonwearable fall detection systems [30] (either based on radar and optical sensors or that use fall mats and bed alarms). As SYG, these emerging technologies warrant further controlled trials in order to assess and compare industrial responses to fall detection needs and their associated costs.

This study has some limitations. All residents were not included: Residents or their legally authorized representatives who did not opt in cited privacy concerns of video recording or that their relative was not yet at fall risk in need of the technology. Regarding variations in TUA, we were able to observe that standalone facilities had higher TUA than memory care units based within larger assisted living facilities, but other unreported factors might account for slower assistance received by fallers. We did not evaluate the cognitive-behavioral disorders in fallers with a formal cognitive assessment, and rates and types of comorbidities related to the long TOG were not measured. In the near future, implementation of randomized trials is needed to compare SYG use versus no use on the potential differences associated with diverse levels of cognitive-behavioral impairment and secondary comorbid outcomes related to falls in parallel to TUA or TOG outcomes. Cost-effectiveness analyses of interventions involving new fall detectors are needed in the field: an economic evaluation would compare costs, including those associated with the new technology and those reduced because unnecessary referrals to the emergency department were avoided.

To conclude, SYG focuses on unmet care needs and efficient new strategies in memory care facilities while enhancing safety in the bedroom and during time zones that represent the highest risk of falls. Unlike other nonmodifiable risk factors in falls [31], TOG can be significantly reduced to improve post-fall prognosis. Automated quantification of TUA and TOG can complement current care practice and support better quality of care for vulnerable individuals. In turn, reducing TOG has the potential to decrease future morbidity, unnecessary referral to the emergency department, and related costs [4,7].

Acknowledgments

The authors wish to thank the care facilities for their support. It was truly inspirational to work with care teams so dedicated to improving the quality of care and who take such pride in the quality of care provided. This research has been supported by the National Institutes of Health (#1R43AG058354) and the National Science Foundation (#1648753).

Conflicts of Interest

EB is an advisory board member of SafelyYou, which is a for-profit entity that provides fall detection and prevention services and has vested stock options in SafelyYou stock. GX was coprincipal investigator on the current study funded by a National Institute of Aging grant and owns stock in SafelyYou. SN is a board member and the chief operations officer of SafelyYou, attends conferences on behalf of SafelyYou, and has vested stock options in SafelyYou as an employee. RS is employed by and owns stock in SafelyYou. PA is a former employee of SafelyYou and owns SafelyYou stock. JJ is a former employee of SafelyYou and owns SafelyYou stock. RH is a former employee of SafelyYou. AB is a former employee of SafelyYou and owns SafelyYou stock. BLM is an advisory board member and owns SafelyYou stock. GN is a board member and the chief executive officer of SafelyYou, attends conferences on behalf of SafelyYou, and owns SafelyYou stock.

References

1. Burns ER, Stevens JA, Lee R. The direct costs of fatal and non-fatal falls among older adults - United States. J Safety Res 2016 Sep;58:99-103. [doi: 10.1016/j.jsr.2016.05.001] [Medline: 27620939]
2. Delbaere K, Kochan NA, Close JC, Menant JC, Sturnieks DL, Brodaty H, et al. Mild Cognitive Impairment as a Predictor of Falls in Community-Dwelling Older People. The American Journal of Geriatric Psychiatry 2012 Oct;20(10):845-853. [doi: 10.1097/jgp.0b013e31824afbc4]
3. van Doorn C, Gruber-Baldini A, Zimmerman S, Hebel J, Port C, Baumgarten M, Epidemiology of Dementia in Nursing Homes Research Group. Dementia as a risk factor for falls and fall injuries among nursing home residents. J Am Geriatr Soc 2003 Sep;51(9):1213-1218. [doi: 10.1046/j.1532-5415.2003.51404.x] [Medline: 12919232]
4. Fleming J, Brayne C, Cambridge City over-75s Cohort (CC75C) study collaboration. Inability to get up after falling, subsequent time on floor, and summoning help: prospective cohort study in people over 90. BMJ 2008 Nov 17;337:a2227 [FREE Full text] [doi: 10.1136/bmj.a2227] [Medline: 19015185]
5. SafelyYou Inc. URL: https://www.safely-you.com/ [accessed 2021-03-29]
6. Bayen E, Jacquemot J, Netscher G, Agrawal P, Tabb Noyce L, Bayen A. Reduction in Fall Rate in Dementia Managed Care Through Video Incident Review: Pilot Study. J Med Internet Res 2017 Oct 17;19(10):e339 [FREE Full text] [doi: 10.2196/jmir.8095] [Medline: 29042342]
7. Xiong GL, Bayen E, Nickels S, Subramaniam R, Agrawal P, Jacquemot J, et al. Real-time video detection of falls in dementia care facility and reduced emergency care. Am J Manag Care 2019 Jul;25(7):314-315 [FREE Full text] [Medline: 31318502]
8. Lamb SE, Jørstad-Stein EC, Hauer K, Becker C, Prevention of Falls Network Europe and Outcomes Consensus Group. Development of a common outcome data set for fall injury prevention trials: the Prevention of Falls Network Europe consensus. J Am Geriatr Soc 2005 Sep;53(9):1618-1622. [doi: 10.1111/j.1532-5415.2005.53455.x] [Medline: 16137297]

9. Davalos-Bichara M, Lin FR, Carey JP, Walston JD, Fairman JE, Schubert MC, et al. Development and validation of a falls-grading scale. J Geriatr Phys Ther 2013;36(2):63-67 [FREE Full text] [doi: 10.1519/JPT.0b013e31825f6777] [Medline: 22810170]
10. Berkeley Human Research Protection Program. UC Berkeley. URL: http://cphs.berkeley.edu/ [accessed 2021-03-30]
11. Rapp K, Becker C, Cameron ID, König HH, Büchele G. Epidemiology of falls in residential aged care: analysis of more than 70,000 falls from residents of bavarian nursing homes. J Am Med Dir Assoc 2012 Mar;13(2):187.e1-187.e6. [doi: 10.1016/j.jamda.2011.06.011] [Medline: 21816682]
12. McCleery J, Cohen D, Sharpley A. Pharmacotherapies for sleep disturbances in dementia. Cochrane Database Syst Rev 2016 Nov 16;11:CD009178 [FREE Full text] [doi: 10.1002/14651858.CD009178.pub3] [Medline: 27851868]
13. Shih Y, Pai M, Huang Y, Wang J. Sundown Syndrome, Sleep Quality, and Walking Among Community-Dwelling People With Alzheimer Disease. J Am Med Dir Assoc 2017 May 01;18(5):396-401. [doi: 10.1016/j.jamda.2016.10.016] [Medline: 28057424]
14. Ohayon MM, Mahowald MW, Dauvilliers Y, Krystal AD, Leger D. Prevalence and comorbidity of nocturnal wandering in the US adult general population. Neurology 2012 May 14;78(20):1583-1589. [doi: 10.1212/wnl.0b013e3182563be5]
15. Pahwa AK, Andy UU, Newman DK, Stambakio H, Schmitz KH, Arya LA. Noctural Enuresis as a Risk Factor for Falls in Older Community Dwelling Women with Urinary Incontinence. J Urol 2016 May;195(5):1512-1516 [FREE Full text] [doi: 10.1016/j.juro.2015.11.046] [Medline: 26626218]
16. Nevitt MC, Cummings SR, Hudes ES. Risk factors for injurious falls: a prospective study. J Gerontol 1991 Sep 01;46(5):M164-M170. [doi: 10.1093/geronj/46.5.m164] [Medline: 1890282]
17. O'Loughlin JL, Robitaille Y, Boivin J, Suissa S. Incidence of and risk factors for falls and injurious falls among the community-dwelling elderly. Am J Epidemiol 1993 Mar 01;137(3):342-354. [doi: 10.1093/oxfordjournals.aje.a116681] [Medline: 8452142]
18. Robinovitch SN, Feldman F, Yang Y, Schonnop R, Leung PM, Sarraf T, et al. Video capture of the circumstances of falls in elderly people residing in long-term care: an observational study. The Lancet 2013 Jan;381(9860):47-54. [doi: 10.1016/s0140-6736(12)61263-x]
19. Ganz DA, Higashi T, Rubenstein L. Monitoring falls in cohort studies of community-dwelling older people: effect of the recall interval. J Am Geriatr Soc 2005 Dec;53(12):2190-2194. [doi: 10.1111/j.1532-5415.2005.00509.x] [Medline: 16398908]
20. Bloch F, Lundy J, Rigaud A. Profile differences of purchasers, non-purchasers, and users and non-users of Personal Emergency Response Systems: Results of a prospective cohort study. Disabil Health J 2017 Oct;10(4):607-610. [doi: 10.1016/j.dhjo.2017.01.008] [Medline: 28159502]
21. Allan LM, Wheatley A, Smith A, Flynn E, Homer T, Robalino S, et al. An intervention to improve outcomes of falls in dementia: the DIFRID mixed-methods feasibility study. Health Technol Assess 2019 Oct;23(59):1-208 [FREE Full text] [doi: 10.3310/hta23590] [Medline: 31661058]
22. Tinetti ME, Liu W, Claus E. Predictors and prognosis of inability to get up after falls among elderly persons. JAMA 1993 Jan 06;269(1):65-70. [Medline: 8416408]
23. Bloch F. Critical falls: why remaining on the ground after a fall can be dangerous, whatever the fall. J Am Geriatr Soc 2012 Jul 12;60(7):1375-1376. [doi: 10.1111/j.1532-5415.2012.04017.x] [Medline: 22788399]
24. Lach HW, Parsons JL. Impact of fear of falling in long term care: an integrative review. J Am Med Dir Assoc 2013 Aug;14(8):573-577. [doi: 10.1016/j.jamda.2013.02.019] [Medline: 23602257]
25. Lach HW. The costs and outcomes of falls: what's a nursing administrator to do? Nurs Adm Q 2010;34(2):147-155. [doi: 10.1097/NAQ.0b013e3181d9185f] [Medline: 20234249]
26. Centre for Clinical Practice at NICE (UK). Falls: Assessment and Prevention of Falls in Older People. London, England: National Institute for Health and Care Excellence; 2013.
27. Panel on Prevention of Falls in Older Persons, American Geriatrics Society and British Geriatrics Society. Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. J Am Geriatr Soc 2011 Jan;59(1):148-157. [doi: 10.1111/j.1532-5415.2010.03234.x] [Medline: 21226685]
28. Silva de Lima AL, Smits T, Darweesh SKL, Valenti G, Milosevic M, Pijl M, et al. Home-based monitoring of falls using wearable sensors in Parkinson's disease. Mov Disord 2020 Jan 26;35(1):109-115 [FREE Full text] [doi: 10.1002/mds.27830] [Medline: 31449705]
29. Aziz O, Park EJ, Mori G, Robinovitch SN. Distinguishing the causes of falls in humans using an array of wearable tri-axial accelerometers. Gait Posture 2014;39(1):506-512. [doi: 10.1016/j.gaitpost.2013.08.034] [Medline: 24148648]
30. Skubic M, Harris BH, Stone E, Ho KC, Bo-Yu S, Rantz M. Testing non-wearable fall detection methods in the homes of older adults. 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 16-20, 2016; Orlando, FL p. 557-560. [doi: 10.1109/embc.2016.7590763]
31. Mangram A, Dzandu J. Why Elderly Patients with Ground Level Falls Die Within 30 Days And Beyond? J Gerontol Geriatr Res 2016;05(02):1. [doi: 10.4172/2167-7182.1000289]

Abbreviations

AI: artificial intelligence
ANOVA: analysis of variance
RTN: real-time notification
SYG: SafelyYou Guardian
TOG: Time On the Ground
TUA: Time Until staff Assistance arrives

Edited by G Eysenbach; submitted 13.01.20; peer-reviewed by N Manley, M Stein, M Immonen, H Wang, J Sosnoff, KS Sczuka; comments to author 28.08.20; revised version received 15.10.20; accepted 11.03.21; published 17.06.21.

Please cite as:
Bayen E, Nickels S, Xiong G, Jacquemot J, Subramaniam R, Agrawal P, Hemraj R, Bayen A, Miller BL, Netscher G
Reduction of Time on the Ground Related to Real-Time Video Detection of Falls in Memory Care Facilities: Observational Study
J Med Internet Res 2021;23(6):e17551
URL: https://www.jmir.org/2021/6/e17551
doi:10.2196/17551
PMID:34137723

Original Paper

A Typology of Existing Machine Learning–Based Predictive Analytic Tools Focused on Reducing Costs and Improving Quality in Health Care: Systematic Search and Content Analysis

Ariadne A Nichol[1], BA; Jason N Batten[1], MA, MD; Meghan C Halley[1], MPH, PhD; Julia K Axelrod[1], BA; Pamela L Sankar[2], PhD; Mildred K Cho[1], PhD

[1]Stanford School of Medicine, Stanford Center for Biomedical Ethics, Stanford, CA, United States

[2]Department of Medical Ethics and Health Policy, Perelman School of Medicine, Philadelphia, PA, United States

Corresponding Author:
Ariadne A Nichol, BA
Stanford School of Medicine
Stanford Center for Biomedical Ethics
1215 Welch Road, Modular A
Stanford, CA, 94305
United States
Phone: 1 650 723 5760
Email: ariadnen@stanford.edu

Abstract

Background: Considerable effort has been devoted to the development of artificial intelligence, including machine learning–based predictive analytics (MLPA) for use in health care settings. The growth of MLPA could be fueled by payment reforms that hold health care organizations responsible for providing high-quality, cost-effective care. Policy analysts, ethicists, and computer scientists have identified unique ethical and regulatory challenges from the use of MLPA in health care. However, little is known about the types of MLPA health care products available on the market today or their stated goals.

Objective: This study aims to better characterize available MLPA health care products, identifying and characterizing claims about products recently or currently in use in US health care settings that are marketed as tools to improve health care efficiency by improving quality of care while reducing costs.

Methods: We conducted systematic database searches of relevant business news and academic research to identify MLPA products for health care efficiency meeting our inclusion and exclusion criteria. We used content analysis to generate MLPA product categories and characterize the organizations marketing the products.

Results: We identified 106 products and characterized them based on publicly available information in terms of the types of predictions made and the size, type, and clinical training of the leadership of the companies marketing them. We identified 5 categories of predictions made by MLPA products based on publicly available product marketing materials: disease onset and progression, treatment, cost and utilization, admissions and readmissions, and decompensation and adverse events.

Conclusions: Our findings provide a foundational reference to inform the analysis of specific ethical and regulatory challenges arising from the use of MLPA to improve health care efficiency.

(J Med Internet Res 2021;23(6):e26391) doi:10.2196/26391

KEYWORDS

machine learning; artificial intelligence; ethics; regulation; health care quality; costs

Introduction

Background

Machine learning–based predictive analytics (MLPA) products are emerging as a strategy for controlling rising health care costs [1]. Advanced statistical analyses have long been used to estimate the likelihood of future health outcomes based on previous events and inform clinical and administrative decisions. MLPA offers a benefit over current approaches because of its ability to draw from larger and more diverse electronic health record (EHR) data sets and potentially draw inferences without human involvement in defining input variables to predict outcomes, with the goal of improving speed and accuracy [2,3]. A particular aspect of health care seen as especially ripe for MLPA application is health care efficiency, which improves patient outcomes while reducing health care costs [4]. By gaining insights from large amounts of clinical information stored in EHR systems, MLPA is being used to identify and direct resources toward patients at higher risk of poor outcomes.

Incentives for health systems to adopt products focused on health care efficiency stem—at least in part—from federal policies and payment structures that encourage value-based care under the Affordable Care Act and value-based purchasing and bundling programs instituted by the Centers for Medicare & Medicaid Services (CMS). For example, the CMS Hospital Readmissions Reduction Program reduces reimbursements to hospitals with excess unplanned 30-day hospital readmissions for certain health conditions [5]. Linking CMS payments to the quality of care creates a financial incentive for health systems to adopt MLPA products that aim to improve health care efficiency.

However, experts recognize that although MLPA could improve the efficiency of delivered care, its use in the health care domain poses distinct ethical challenges because of its lack of transparency, continuous adaptation without human intervention, and its potential for systematic error leading to unfair decisions or actions [6-8]. These challenges have been demonstrated by high-profile cases, such as the predictive risk-stratification algorithm developed by Optum, which resulted in significant racial bias against Black patients when health costs were used as a proxy measure of health needs [9]. Obtaining, sharing, and handling the sensitive data necessary for MLPA in health care also raises privacy concerns. In another high-profile case, technology giant Google was sued for alleged violation of the Health Insurance Portability and Accountability Act in its handling of patient records from the University of Chicago Medical Center for the development of predictive artificial intelligence tools [10]. More recently, Google's partnership with Ascension, one of the largest private, faith-based health care systems in the United States, came under investigation after revealing that the tech company had obtained protected health information without patient consent [11,12]. Many of the ethical challenges mentioned, such as privacy concerns, are not unique to the utilization of MLPA and are relevant to other uses of advanced statistical analyses implemented in health care. However, continuous self-learning and the lack of transparency in MLPA algorithms are two unique aspects of the techniques that make it difficult to evaluate the models and understand how decisions are made.

These cases also highlight the importance and challenges of oversight of these complex software products. Unlike drugs and medical devices that the Food and Drug Administration (FDA) typically regulates, MLPA-based products are constantly and inherently mutable, complicating the definition of the final product. The US FDA is actively testing a regulatory framework for software as a medical device through a precertification pilot program. The framework shifts the emphasis away from the evaluation of completed products to the evaluation of processes that demonstrate a "culture of quality and organizational excellence" [13,14]. It is unclear whether or how such a framework applies to MLPA products to improve health care efficiency or what features constitute a *culture of quality and organizational excellence* capable of facilitating the development of safe and effective products. This ambiguity is because of, at least in part, a lack of systematic information on the characteristics of organizations that develop such MLPA, which is essential to understanding their potential ability to self-regulate. The diversity of expertise required to develop MLPA for health care, including computer science, software engineering, and medicine, suggests that teams brought together to develop and implement particular products will include members who are unfamiliar with the norms and culture of biomedical research and development and clinical practice [15,16]. The types of organizations developing MLPA products and the types of expertise at the organizations are largely unknown, precluding analysis of the alignment of interests and expertise with the needs of patients and health care providers [17].

Objective

The main objective of this study is to map the landscape of currently available MLPA products marketed with the aim of improving health care efficiency. The study also seeks to characterize organizations developing these MLPA products, with the subsequent goal of identifying relevant ethical, regulatory, and policy implications.

Methods

Search Strategy

We sought to identify MLPA products based on publicly available marketing information. To identify these products, we assessed 4 databases: LexisNexis, PubMed, Web of Knowledge, and Indeed.com. PubMed references frequently omitted necessary details to judge a product's current use, and many of the results were duplicative with Web of Knowledge results. On this basis, we eliminated PubMed and conducted our research using the other 3 databases. LexisNexis searches returned the highest number of nonduplicative results. Indeed.com (the world's largest job listing website) and Web of Knowledge were used because they returned additional nonduplicative results. Search terms such as "hospitals," "health care organizations," "machine learning," and "predictive analytics" were used (see Multimedia Appendix 1 for further details). Search terms were tested, reviewed, and refined to maximize the number of returned MLPA products that met our inclusion criteria. LexisNexis search parameters included the United States and English language, and a date range from April 1, 2015, to February 1, 2019, selected to capture efforts launched likely in response to April 2015 American Recovery and Reinvestment Act-mandated changes to CMS reimbursement policies.

Eligibility Criteria and Screening Process

We first removed all duplicates and any results that did not mention specific products (ie, congressional transcripts). For the identified products, we conducted additional targeted searches as needed to elucidate whether specific products met the eligibility criteria. The final list of eligible products for which the marketing materials were identified made the following claims: (1) the MLPA product made health care–related predictions, (2) the product primarily aimed to improve health care quality and reduce costs (ie, improve health care efficiency), (3) the product used EHR-sourced data, and (4) the product had been implemented by an identifiable US health system or provider, and possibly, though not necessarily, utilized on a routine basis. In addition, we excluded products if, based on marketing language, they (1) lacked a predictive element, (2) were not directly related to improving the quality of delivered care (eg, managing appointment schedules), or (3) solely used patient data that were not EHR-sourced (eg, data from a wearable device).

Data Extraction

For all remaining products, we used the product website to collect additional information about the product characteristics and the organization that developed it. Characteristics included health care partners using the product, sources of data used to create and train the MLPA algorithms, and the type and size of the organization marketing the product. We also characterized the companies by the number of employees, the type of business, and whether the chief executives or board members had a clinical degree, including doctor of medicine, registered nurse, or other.

Data Analysis

We used content analysis to generate MLPA product categories based on the type of prediction made [18]. To do this, we first generated initial codes from the verbatim marketing language available on the organization websites. If a single product made multiple predictions, we applied multiple codes to capture all the product predictions, and 2 researchers independently coded 50% of the products. All discrepancies were resolved through discussion with the full research team. A single researcher coded the remaining products. After all MLPA products received initial codes, we grouped the initial codes into overarching categories based on the prediction type. We counted the total number of products in each category as well as membership in multiple categories.

Results

Search Results and Characterization of Companies

From 1288 articles and other sources, we found 106 MLPA products developed by 96 companies that met our inclusion and exclusion criteria (Figure 1). The products and characteristics of the companies are listed in Table 1.

Figure 1. Identification of machine learning–based predictive analytics products.

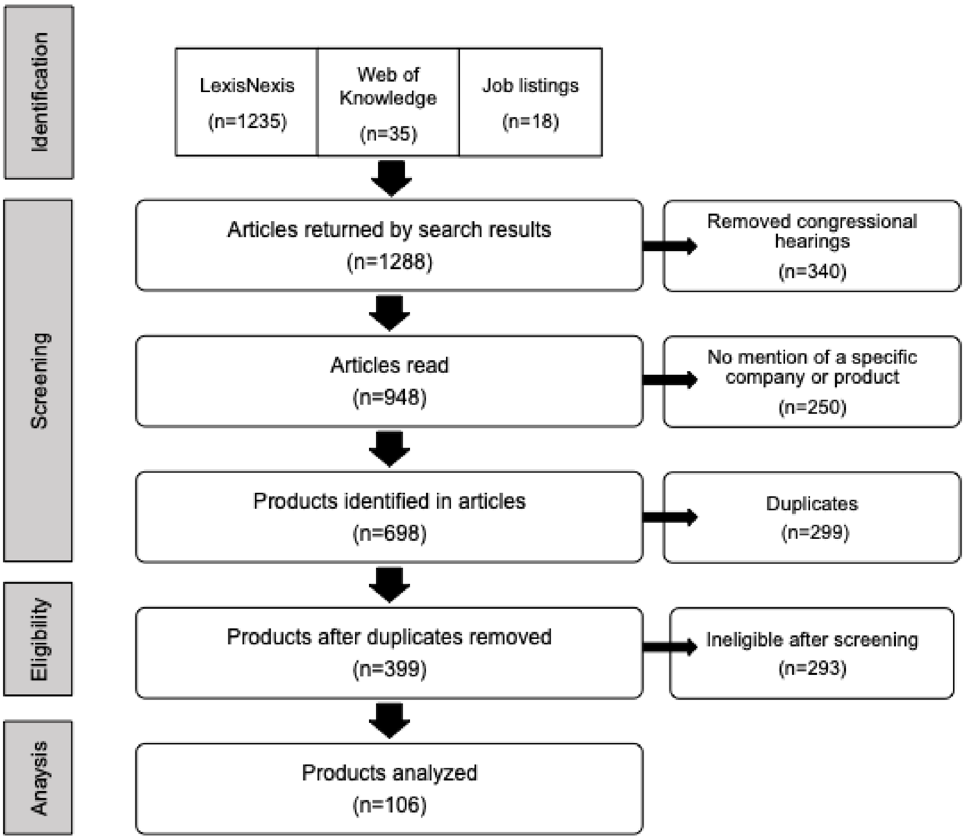


Table 1. Characteristics of companies developing machine learning–based predictive analytics products (N=96).

Characteristics and categories	Values, n (%)
Organization size	
Small (1-50 employees)	34 (35)
Medium (51-1000 employees)	25 (26)
Large (more than 1000 employees)	37 (39)
Organization type	
Computer software company—health care	68 (71)
Computer software company—general	14 (15)
Health insurer	6 (6)
Provider (hospital or health system)	8 (8)
CEO[a] with a clinical degree	
Yes	15 (16)
No	81 (84)
C-suite or board member with a clinical degree	
Yes	62 (65)
No	34 (35)

[a] CEO: chief executive officer.

Many organizations did not meet the inclusion criteria because their products were not yet implemented by an identifiable health system or provider. Other organizations were similarly ineligible because the marketing language did not claim that their product used MLPA for predicting how to reduce cost and improve quality of care. Of the organizations, 92% (88/96) developed 1 product that met the inclusion criteria, whereas 8% (8/96) had more than one product. Companies were broadly distributed in terms of size. The vast majority 85% (82/96) were computer software companies, of which 83% (68/82) specialized in health care–related products. Of the MLPA developers, 15% (14/96) were health insurers, hospitals, or health systems.

Although chief executive officers (CEOs) of 84% (81/96) of companies did not have a clinical degree, 65% (62/96) listed a C-suite or board member who did. Of the software companies specializing in health care, 16% (11/68) had a clinician CEO, and 72% (49/68) had a clinician C-suite or board member. Computer software companies specializing in health care made up 94% (32/34) of small organizations with 50 employees or less. None of the large general computer software companies had a clinician as CEO, 75% (9/12) had a chief medical officer, and 8% (1/12) had a clinician C-suite or board member. All providers (hospitals or health systems) were large organizations with more than 1000 employees. Of the providers, 50% (4/8) had a clinician as CEO, and all providers had a clinician C-suite or board member.

Classification of MLPA Products

Overview

We identified 5 categories of predictions made by MLPA products based on the publicly available product marketing materials: disease onset and progression, treatment, cost and utilization, admissions and readmissions, and decompensation and adverse events (Table 2).

Of the products, 67% (71/106) were assigned to more than one category. A full list of products and their assigned categories can be found in Multimedia Appendix 2. Here, we describe the categories qualitatively and describe a typical product from each category.

Table 2. Categories of predictions made by MLPA products.

MLPA[a] prediction category[b]	Examples of specific predictions	Example quotes from product descriptions provided by developers
Disease onset and progression predictions (n=62)	Patient outcome; unspecified diseases; chronic illnesses; specified diseases; mortality; comorbidities	• "Enables early prediction of disease onset." • "Clinicians can now see red flags for admitted patients at elevated risk of mortality three to five days in advance."
Treatment predictions (n=48)	Best course of treatment; candidates for palliative care or hospice; untreated or undertreated individuals (often referred to as *gaps in care*); expected recovery trajectory; type of care required; best medication or drug efficacy; patients at risk of receiving unnecessary clinical care (visits, tests, procedures, or antibiotics); next steps of medical care that a physician would order	• "Identify members earlier in their disease progression who are likely going to be overmedicalized during the last 6-12 months of life." • "Helps clinicians make data-driven decisions about a patient's care plan."
Cost and utilization predictions (n=38)	High-cost members of a population; high utilizers in a population; risk stratification; cost of caring for a specific patient; Medicare's predicted risk	• "Predict health care cost for individuals for customer specified time periods."
Decompensation and adverse events predictions (n=34)	Hypotensive event; sepsis; hemodynamic instability; inpatient or outpatient decompensation; postoperative complications or surgical site infections; risk of adverse event; adverse medication reactions; hospital-acquired infection; hospital-acquired pressure injury	• "Identify patients at risk of surgical site infection." • "A respiratory failure detection algorithm...can highlight patients at a higher risk of prolonged ventilation up to 48 hours before onset."
Admissions and readmissions predictions (n=33)	Readmission risk; avoidable hospital admission or readmission or ED[c] use; unplanned ICU[d] admission or readmission; ED presentation volume; hospitalization; patient flow; length of stay or risk of an extended length of stay; discharge date; disposition at the end of hospitalization	• "Predicted output is the % chance that the patient will not return/be readmitted." • "Using only six vital signs and patient age, our machine learning tool more accurately predicted down-transfer success."

[a] MLPA: machine learning–based predictive analytics.

[b] Totals do not add up to 106 because categories are not mutually exclusive.

[c] ED: emergency department.

[d] ICU: intensive care unit.

Disease Onset and Progression

A total of 62 products were used to predict the disease onset and progression (see Textbox 1 for an example product). The marketing language for some of these products did not specify particular conditions or diseases. For products that did specify diseases or health states, 22 products were identified as predicting the onset and development of diabetes, cancer, or cardiovascular conditions. In addition, 5 of those 22 products predicted more than one of the 3 listed conditions. Nearly half of these products also explicitly performed cost and utilization prediction instead of simply providing data that could be used to reduce cost.

Textbox 1. Example of a disease onset and progression product.

Medictiv by CitiusTech

- CitiusTech is a large private health care information technology company. Medictiv is a statistical analysis tool advertised as having machine learning capabilities to analyze longitudinal electronic health record–sourced data to predict the onset and progression of various unspecified diseases. Medictiv also advertises specific use cases for chronic kidney disease (CKD) and diabetes. For CKD, Medictiv uses longitudinal patient and laboratory data to predict disease progression risk for CKD stage 3 patients. For diabetes, Medictiv uses data available within 72 hours of admission, including laboratory results, demographic data, comorbidities, and health insurance claims to predict patients' length of stay, risk of readmission, and risk stratification [19].

Treatment

A total of 48 products made predictions related to patient treatment (see Textbox 2 for an example product). The most common type of prediction was identifying patients with *care gaps* or who were *untreated* or *undertreated*. The available marketing language does not specify the meaning of these terms. Nearly all products identifying care gaps explicitly mentioned performing a cost and utilization prediction. In the treatment category, other products made predictions meant to aid clinicians in therapeutic decision making (eg, identifying the best medication to use, predicting adverse reactions, predicting drug interactions, and predicting unnecessary antibiotic use). A large proportion (29/48, 60%) of the products in the treatment category were also categorized as predicting disease onset and progression.

Textbox 2. Example of a treatment product.

Identifi by Evolent Health

- Identifi is Evolent Health's value-based care product, which aims to reduce costs and improve the quality of delivered care. Identifi's machine learning–based predictive analytics algorithms use clinical, social, and administrative data to predict the best course of treatment for a patient and identify gaps in a patient's care. They also make predictions about patient outcomes, risk of readmission, and risk stratification. Evolent Health is a public health care company with between 1000 and 5000 employees. It advertises Identifi to providers and health plans [20].

Cost and Utilization

This category comprises products whose MLPA algorithms predict the cost or utilization of health care (n=38; see Textbox 3 for an example product). As one of our inclusion criteria was a focus on improving health care efficiency, all products in our sample had the goal of reducing costs. However, only products that aimed to reduce costs by making explicit predictions about costs and utilization were included in this category. As there were numerous products that used MLPA primarily to predict admissions or readmissions (n=33), they were assigned their own category (described below and not included in the n=38 of the cost and utilization category). In the cost and utilization category, financial risk stratification of patient populations was the most common MLPA use, with a wide margin. Other common use cases were predicting which patients would be high cost or become high utilizers and predicting Medicare's predicted risk. Half of the products in this category also fell into the categories of disease onset, progression predictions, and treatment predictions.

Textbox 3. Example of a cost and utilization product.

Waystar Platform by Waystar

- Waystar uses social determinants of health, along with hospital and consumer data, to stratify the patient population according to risk and cost [21]. The company also helps with revenue integrity by identifying incorrectly coded and undercoded claims to help providers maximize revenue. Waystar is a medium-sized private information technology company with 500-1000 employees.

Decompensation and Adverse Events

The products in this category (n=34) were designed to act as early warning systems for the occurrence of adverse events or decompensations (see Textbox 4 for an example product). We grouped decompensation and adverse event predictions together when we defined our prediction categories because of the frequent overlap in the clinical application of the products. Algorithms typically use vital signs combined with EHR data to closely monitor in-patient populations and alert care teams for decompensation. Products alerting for general inpatient decompensation were the most common in this category (n=18), followed by early warning systems for sepsis (n=14). Monitoring for an outpatient decline, hospital-acquired infections, and postoperative complications were also quite common. The decompensation and adverse events category had the least overlap with other categories.

Textbox 4. Example of a decompensation and adverse events product.

InSight by Dascena

- Dascena's InSight is a paradigmatic application of machine learning–based predictive analytics used to provide an early warning of an adverse event. Dascena is a small, private company with less than 50 employees. The InSight algorithm warns of sepsis onset using vital sign data located in patients' electronic health records, which is typical of products in this category. InSight provides physicians with real-time alerts and boasts its ability to forecast a patient's condition 4 hours in the future [22].

Admissions and Readmissions

In this category (n=33), predicting the risk of readmission was the most common application (n=21), where the marketing language had to explicitly state *predicting risk of readmission* as a use case of the product (see Textbox 5 for an example product). Other admissions predicted length of stay and intensive care unit occupancy. More than half of the products predicting admissions and readmissions fell into the disease onset and progression category; overlap with the treatment category was also fairly high (8/33, 24%).

Textbox 5. Example of an admissions and readmissions product.

Conduent's Midas Readmission Penalty Forecaster

- Midas Readmission Penalty Forecaster is a common product developed in response to Centers for Medicare & Medicaid Services Hospital Readmissions Reduction Program [23]. Conduent's white pages market the product as a web-based tool to forecast 30-day unplanned readmissions to help health care organizations predict penalties and adjust their care delivery. Midas Readmission Penalty Forecaster estimates total readmissions, excess readmissions, and financial penalties for 6 Centers for Medicare & Medicaid Services cohorts: patients diagnosed with acute myocardial infarction, heart failure, pneumonia, or chronic obstructive pulmonary disease or patients receiving a coronary artery bypass graft, total hip arthroplasty, or total knee arthroplasty. This product was developed by Conduent, an information technology company with more than 10,000 employees working in health care and 20 other industries, ranging from insurance and government to casinos, oil, and gas.

Discussion

Principal Findings

Our results provide an overview of the emerging MLPA applied to improve health care efficiency and provide a systematic categorization of actual applications of this technology in patient care. The products identified as being currently in use are predominantly marketed by computer software companies. Our results also provide a systematic framework for mapping the characteristics of organizations operating in the field of MLPA in health care and the products they produce, based on the specific predictions that these products are intended to provide in a health care setting.

The potential for MLPA to transform health care has generated much anticipation to the possibilities for this technology to improve health care quality and reduce costs. Bates et al [2] previously forecasted 6 likely targets of predictive analytics in health care: high-cost patients, readmissions, decompensation, adverse events, triage, and diseases affecting multiple organ systems. Our results suggest that many of these uses materialized in the markets. For example, our results confirmed a large market presence of MLPA products that aim to predict hospital readmission within 30 days of initial discharge as well as decompensation and adverse events. In addition, among the MLPA products categorized as predicting disease onset and progression within our framework, diabetes, cancer, or cardiovascular conditions were the most common—conditions all affecting multiple organ systems. Furthermore, our categories focused on prediction of cost and utilization as well as on prediction of treatment; both have close ties to the previously forecasted focus of predictive analytics on high-cost patients and triage in health care. However, our framework also goes beyond these original predictions by providing a systematic, evidence-based approach to mapping the field of MLPA products in health care organized around the specific predictions provided by these products instead of the intended use or target population.

Our results also suggest that MLPA products are increasingly being used in response to CMS reimbursement policies. The readmissions predictions may reflect a response to the recent CMS Hospital Readmissions Reduction Program, which reduces payments to hospitals with excessive numbers of readmissions [5]. The focus on diabetes, cancer, or cardiovascular conditions in MLPA products identified in our analysis maps directly to conditions subjected to bundled payments under the CMS's Bundled Payments for Care Improvement initiative [24]. In addition, products marketed as predicting high-cost patients (while also identifying some additional applications, such as predicting Medicare's predicted risk score) likely emerged, at least in part, in response to Medicare's reimbursement policies transitioning from fee-for-service to risk-adjusted fixed payments per episode of care [25]. Although the use of MLPA to respond to shifting reimbursement policies is perhaps unsurprising, it also raises questions about the alignment of these financial incentives with the goal of improving patient care. These goals address critically important needs of the health care system in the United States, but trying to meet them can raise ethical issues. Improving care quality and outcomes without increasing costs poses a myriad of challenges. Thus, when efficiency is improved by reducing costs, there are concerns that quality of care has been negatively affected. Although the aims of improving quality and cost are ideally aligned with the stated goals of these MLPA products, it is difficult to know whether this is indeed the case when employing MLPA products without further information on how the underlying models are developed and implemented in the clinical setting. Moreover, many MLPA algorithms have not been rigorously tested, and little is known about their comparison with other predictive analytics or clinical judgment. Evaluation of MLPA algorithms is particularly difficult given the opacity of the models and their *black box* nature.

Our results also provide an essential framework for considering various approaches to regulation in this diverse and rapidly changing marketplace. The FDA is currently developing a framework that incorporates the level of risk to the patient in its review process. Having a systematic framework of categories that may reflect varying degrees or types of risk to patients (eg, treatment recommendations vs prediction of health care costs) may therefore be important. Traditionally, software products have not been subjected to the level of regulatory scrutiny applied to drugs or medical devices, nor has the technology sector established processes for identifying or evaluating ethical issues that may arise from their products. Developing an effective regulatory framework requires an understanding of various stakeholders and organizations involved in this marketplace, potential sources of conflict, and the resources necessary for success. In examining MLPA products, which inherently change and adapt as they incorporate new data, regulators may need to consider the extent to which business requirements—including production schedules, fundraising, and profit goals—are aligned with the design process.

In addition, further examination is needed regarding the role of clinical expertise within these companies in light of the FDA's self-regulation approach in evaluating companies based on *a culture of quality and organizational excellence*. There is a relative dearth of clinical training among CEOs and others in company leadership. Of the organizations we identified, only 1 in 6 was led by a clinical degree-holding CEO, and more than a third did not have a clinician in the C-suite or on their board of trustees. Although clinicians may be involved in different roles, they are underrepresented in the highest leadership positions, which may have implications for the level of awareness that a company has of the norms and culture of biomedical research and clinical practice. The influence of business requirements and expertise may also vary depending on the size of the company: although small health care technology startups might be under more significant financial and time pressures with the need to raise venture capital, larger companies likely have more resources to draw from. However, our analysis suggests that large companies are also less likely to specialize in health care software or technology and less likely to have a clinician in a leadership position than small health care technology startups. More research is needed to determine the extent to which factors such as company size, business requirements, and clinical expertise influence the design and implementation of MLPA in health care and their potential importance in designing regulatory frameworks.

Our study has several limitations. Our results are limited by our reliance on publicly available web-based information, such as product websites, press releases, and health system websites. Products developed by nonprofit health systems, academic institutions, or large insurers may not have been readily identifiable, as their products are often not marketed externally. Therefore, we are less likely to have identified products developed by a health system or health insurer that are not sold for use in other systems. Another limitation is that the predictions were categorized based on the marketing language used by the companies to describe their own products, so the actual extent to which these products do what they are marketed to do remains unclear. In addition, we do not know how often the tools are used by the health care system where they are implemented. Some may be used frequently and others rarely.

Conclusions

There is a rapidly emerging set of products that utilize MLPA with the dual goals of improving health care and addressing cost containment. These goals address critically important needs of the health care system in the United States. Improving care quality and outcomes is not necessarily at odds with lowering costs. There is an underlying ethical tension, however, when health care efficiency is improved by reducing cost with possible negative effects on quality. How MLPA developers perceive these trade-offs and whether reliance on such tools may exacerbate discrimination based on underlying biases is difficult to assess using currently available data. The significant role of the software and technology companies, which might have little experience in understanding clinical care, using health data, or applying medical ethics or law, suggests that regulatory approaches that rely on self-regulation and organizational culture may be challenging for the evaluation of MLPA products. More research on the process of developing these novel tools is needed to further assess the implications for policy and regulation.

Acknowledgments

This study was supported by a grant from the Greenwall Foundation, *Identifying potential barriers to and enablers of the development of ethical machine learning for health care*, and National Institutes of Health grants 1R01HG010476 and T32 HG008953.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms.
[DOC File , 31 KB - jmir_v23i6e26391_app1.doc]

Multimedia Appendix 2
Categorization of machine learning–based predictive analytics products.
[DOC File , 143 KB - jmir_v23i6e26391_app2.doc]

References

1. National Health Expenditure Projections 2018-2027. Centers for Medicare and Medicaid Services. 2018. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/ForecastSummary.pdf [accessed 2020-09-04]
2. Bates D, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014 Jul;33(7):1123-1131 [FREE Full text] [doi: 10.1377/hlthaff.2014.0041] [Medline: 25006137]
3. Koh H, Tan G. Data mining applications in healthcare. J Healthc Inf Manag 2005;19(2):64-72. [Medline: 15869215]
4. Atasoy H, Greenwood B, McCullough J. The digitization of patient care: a review of the effects of electronic health records on health care quality and utilization. Annu Rev Public Health 2019 Apr 01;40:487-500 [FREE Full text] [doi: 10.1146/annurev-publhealth-040218-044206] [Medline: 30566385]
5. Hospital Readmissions Reduction Program (HRRP). Centers for Medicare and Medicaid Services. 2018. URL: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program [accessed 2020-09-04]
6. Artificial intelligence (AI) in healthcare and research. Nuffield Council on Bioethics. 2018. URL: https://www.nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research.pdf [accessed 2021-06-03]
7. Ferryman K, Pitcan M. Fairness in precision medicine. Data&Society. 2018. URL: https://datasociety.net/wp-content/uploads/2018/02/Data.Society.Fairness.In_.Precision.Medicine.Feb2018.FINAL-2.26.18.pdf [accessed 2020-09-04]

8. Cohen I, Amarasingham R, Shah A, Xie B, Lo B. The legal and ethical concerns that arise from using complex predictive analytics in health care. Health Aff (Millwood) 2014 Jul;33(7):1139-1147 [FREE Full text] [doi: 10.1377/hlthaff.2014.0048] [Medline: 25006139]
9. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019 Oct 25;366(6464):447-453 [FREE Full text] [doi: 10.1126/science.aax2342] [Medline: 31649194]
10. Wakabayashi D. Google and the University of Chicago are sued over data sharing. The New York Times. 2019. URL: https://www.nytimes.com/2019/06/26/technology/google-university-chicago-data-sharing-lawsuit.html [accessed 2021-06-03]
11. Copeland R, Needleman S. Google's 'Project Nightingale' triggers federal inquiry. The Wall Street Journal. 2019. URL: https://www.wsj.com/articles/behind-googles-project-nightingale-a-health-data-gold-mine-of-50-million-patients-11573571867 [accessed 2020-09-04]
12. Copeland R. Google's 'Project Nightingale' gathers personal health data on millions of Americans. The Wall Street Journal. 2019. URL: https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790 [accessed 2021-06-03]
13. Digital health software precertification (Pre-Cert) program: participate in 2019 test plan. U.S. Food & Drug Administration (FDA). 2020. URL: https://www.fda.gov/medical-devices/digital-health-software-precertification-pre-cert-program/digital-health-software-precertification-pre-cert-program-participate-2019-test-plan [accessed 2020-09-04]
14. Digital health innovation action plan. U.S. Food & Drug Administration (FDA). URL: https://www.fda.gov/media/106331/download [accessed 2020-09-04]
15. The social and economic implications of artificial intelligence technologies in the near-term: a summary of the AI Now public symposium, hosted by the White House and New York University's Information Law Institute. The AI Now Report. 2016. URL: https://ainowinstitute.org/AI_Now_2016_Report.pdf [accessed 2020-09-04]
16. Passi S, Barocas S. Problem formulation and fairness. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019 Presented at: FAT* '19: Conference on Fairness, Accountability, and Transparency; January 29-31, 2019; Atlanta GA USA p. 39-48 URL: http://dl.acm.org/citation.cfm?doid=3287560.3287567 [doi: 10.1145/3287560.3287567]
17. Evans JM. The changing ethics of health care. Caring Ages 2016 Jul;17(7):12 [FREE Full text] [doi: 10.1016/j.carage.2016.06.012]
18. Hsieh H, Shannon S. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288 [FREE Full text] [doi: 10.1177/1049732305276687] [Medline: 16204405]
19. Case study: predicting LOS for patients with diabetes. CitiusTech. URL: https://web.citiustech.com/platforms/medictiv [accessed 2020-09-04]
20. Manage your entire value business from a single, integrated platform. Evolent Health Services. URL: https://www.evolenthealth.com/value-based-technology-platform [accessed 2020-09-04]
21. Patient insights: gain a deeper understanding of health. Waystar. URL: https://www.waystar.com/our-platform/patient-insights/ [accessed 2020-09-04]
22. Shimabukuro D, Barton C, Feldman M, Mataraso S, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res 2017;4(1):e000234 [FREE Full text] [doi: 10.1136/bmjresp-2017-000234] [Medline: 29435343]
23. Midas readmission penalty forecaster. Conduent. 2017. URL: https://downloads.conduent.com/content/usa/en/brochure/midas-readmission-penalty-forecaster.pdf [accessed 2020-09-04]
24. Bundled payments for care improvement (BPCI) initiative: general information. Centers for Medicare & Medicaid Services. URL: https://innovation.cms.gov/innovation-models/bundled-payments [accessed 2020-09-04]
25. Acute care hospital inpatient prospective payment system. Centers for Medicare and Medicaid Services. 2019. URL: https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/AcutePaymtSysfctsht.pdf [accessed 2020-09-04]

Abbreviations

CEO: chief executive officer
CMS: Centers for Medicare & Medicaid Services
EHR: electronic health record
FDA: Food and Drug Administration
MLPA: machine learning–based predictive analytics

Edited by R Kukafka, G Eysenbach; submitted 09.12.20; peer-reviewed by A Mehrotra, G Murray; comments to author 27.02.21; revised version received 12.03.21; accepted 17.05.21; published 22.06.21.

Please cite as:
Nichol AA, Batten JN, Halley MC, Axelrod JK, Sankar PL, Cho MK
A Typology of Existing Machine Learning–Based Predictive Analytic Tools Focused on Reducing Costs and Improving Quality in Health Care: Systematic Search and Content Analysis
J Med Internet Res 2021;23(6):e26391
URL: https://www.jmir.org/2021/6/e26391
doi:10.2196/26391
PMID:34156338

Original Paper

Discovery of Depression-Associated Factors From a Nationwide Population-Based Survey: Epidemiological Study Using Machine Learning and Network Analysis

Sang Min Nam[1*], MD, PhD; Thomas A Peterson[2*], PhD; Kyoung Yul Seo[3], MD, PhD; Hyun Wook Han[4], MD, PhD; Jee In Kang[5], MD, PhD

[1]Department of Ophthalmology, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea

[2]UCSF REACH Informatics Core, Department of Orthopaedic Surgery, Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA, United States

[3]Department of Ophthalmology, Institute of Vision Research, Eye and Ear Hospital, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

[4]Department of Biomedical Informatics, CHA University School of Medicine, CHA University, Seongnam, Republic of Korea

[5]Department of Psychiatry, Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jee In Kang, MD, PhD
Department of Psychiatry
Institute of Behavioral Science in Medicine
Yonsei University College of Medicine
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 1620
Fax: 82 2 313 0891
Email: jeeinkang@yuhs.ac

Abstract

Background: In epidemiological studies, finding the best subset of factors is challenging when the number of explanatory variables is large.

Objective: Our study had two aims. First, we aimed to identify essential depression-associated factors using the extreme gradient boosting (XGBoost) machine learning algorithm from big survey data (the Korea National Health and Nutrition Examination Survey, 2012-2016). Second, we aimed to achieve a comprehensive understanding of multifactorial features in depression using network analysis.

Methods: An XGBoost model was trained and tested to classify “current depression” and “no lifetime depression” for a data set of 120 variables for 12,596 cases. The optimal XGBoost hyperparameters were set by an automated machine learning tool (TPOT), and a high-performance sparse model was obtained by feature selection using the feature importance value of XGBoost. We performed statistical tests on the model and nonmodel factors using survey-weighted multiple logistic regression and drew a correlation network among factors. We also adopted statistical tests for the confounder or interaction effect of selected risk factors when it was suspected on the network.

Results: The XGBoost-derived depression model consisted of 18 factors with an area under the weighted receiver operating characteristic curve of 0.86. Two nonmodel factors could be found using the model factors, and the factors were classified into direct ($P<.05$) and indirect ($P\geq.05$), according to the statistical significance of the association with depression. Perceived stress and asthma were the most remarkable risk factors, and urine specific gravity was a novel protective factor. The depression-factor network showed clusters of socioeconomic status and quality of life factors and suggested that educational level and sex might be predisposing factors. Indirect factors (eg, diabetes, hypercholesterolemia, and smoking) were involved in confounding or interaction effects of direct factors. Triglyceride level was a confounder of hypercholesterolemia and diabetes, smoking had a significant risk in females, and weight gain was associated with depression involving diabetes.

Conclusions: XGBoost and network analysis were useful to discover depression-related factors and their relationships and can be applied to epidemiological studies using big survey data.

(J Med Internet Res 2021;23(6):e27344) doi:10.2196/27344

KEYWORDS

depression; epidemiology; machine learning; network; prediction model; XGBoost

Introduction

Importance of Depression

Depression is a common debilitating psychiatric condition characterized by a low-spirited mood, loss of interest, and a range of emotional, cognitive, physical, and behavioral symptoms. It has a high global disease burden and had been projected to become the second most common cause of disability-adjusted life years worldwide by 2020 [1]. Thus, the determination of risk factors causing depression could have important implications in its prevention and intervention efforts by reducing modifiable risk factors. However, proper prevention and treatment of depression have been difficult owing to heterogeneity in the etiology and pathophysiology of depression [2].

Conventional Modeling for Depression

Given the complex biological, psychological, and sociocultural factors underlying the pathogenesis of depression, an integrated model with confounder adjustment may provide a better understanding and multifaceted individualized approach for depression. In a typical study, survey-weighted logistic regression is used to identify depression-associated factors. A simple regression model for each candidate factor is built to adjust for age and sex. Then, a complex model is presented by adding more covariates to control potential confounders [3].

Principal Factor Identification Problem in Conventional Modeling

A confounder has associations with both exposure and disease but is not in the causal pathway between exposure and outcome [4]. Confounding is a mixing of effects that obscures the real effect of exposure [4]. We might fit a regression model on all the measured potential confounders and minimize the confounding risk [5]. However, too many variables can cause multicollinearity or redundancy and make a regression model unstable with a high coefficient variance [5,6]. In contrast, variable selection may cause the exclusion of essential confounders [5]. Identification of the best subset of variables is challenging when the number of explanatory variables is large or when multicollinearity is present within the data [7].

Benefits of Machine Learning–Based Modeling for Big Survey Data

For big survey data with numerous variables, the machine learning model can detect principle factors for a condition by attempting to maximize its predicting performance. During the training, dimensionality reduction (which reduces the number of features) is essential because many features may cause overfitting, and the generated model may not be generalized appropriately [8]. Moreover, feature selection can resolve multicollinearity by removing redundant features from a group of highly correlated features [8]. Regularization is a more advanced machine learning algorithm that reduces model complexity and improves generalizability [8]. A popular regularization method is LASSO (L1 regularization; least absolute shrinkage and selection operator) [8], also called a shrinkage method, which can be applied to collinear confounders and has shown advantages over conventional methods [5].

Advantages of the Extreme Gradient Boosting Algorithm and Strategies of This Study

We speculate that extreme gradient boosting (XGBoost) can help select an optimal subset of essential variables from a large volume of survey data. XGBoost is an advanced implementation of a gradient-boosting decision-tree algorithm. XGBoost has been used in a few studies to predict or screen depression [9-11]. XGBoost is advantageous because of its high speed and performance, making it dominant in applied machine learning for structured data. XGBoost also offers regularized gradient boosting and feature importance scores using a trained predictive model, which can be used for feature selection [12].

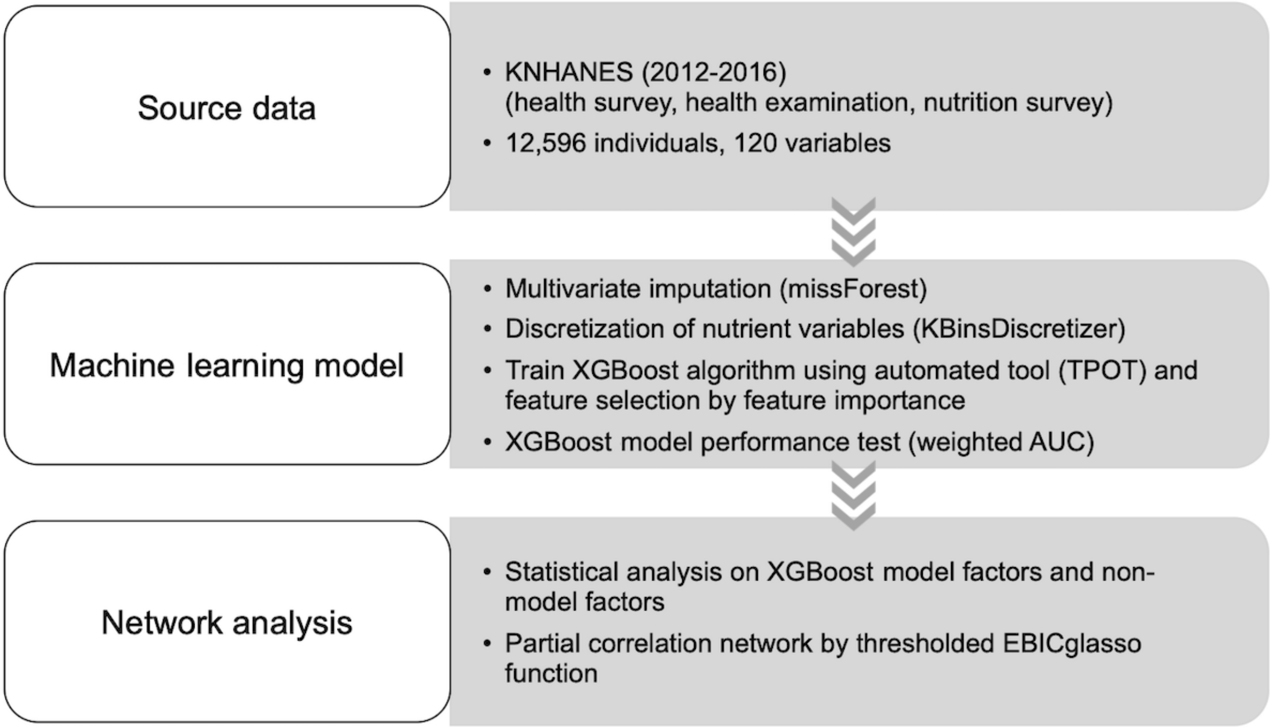
However, XGBoost has many hyperparameters that should be manually specified, and optimal parameter tuning can be challenging. To solve this problem, we have chosen an automated machine learning tool, the tree-based pipeline optimization tool (TPOT), along with genetic programming [13]. To the best of our knowledge, this study is the first attempt at an XGBoost-supported epidemiologic investigation regarding depression in the general population using national survey data.

Methods

Overview of Survey Data

The overall study flowchart is provided in Figure 1. We used survey data obtained from 12,596 Koreans (aged 19-64 years) who participated in the Korea National Health and Nutrition Examination Survey (KNHANES, 2012-2016), which is an annual survey conducted by the Korea Centers for Disease Control and Prevention (KCDC) in the Republic of Korea, wherein population-wide health and nutritional statuses are assessed [14]. The maximum age was set to 64 years because the food frequency survey was conducted for those aged 19 to 64 years. The design and methods of the KNHANES and the data resource profile are available in previous reports [14-16]. KNHANES V-3 (2012), KNHANES VI (2013-2015), and KNHANES VII-1 (2016) were approved by the KCDC research ethics committee (2012-01EXP-01-2C, 2013-07CON-03-4C, 2013-12EXP-03-5C, and 2015-01-02-6C), and written informed consent was obtained from all the subjects.

Figure 1. Flow diagram of the study. AUC: area under the curve; EBICglasso: extended Bayesian information criterium graphical lasso; KNHANES: Korea National Health and Nutrition Examination Survey.



In the annual KNHANES (2012-2016) following a multistage clustered probability design, 192 primary sampling units (PSUs) were generated from approximately 200,000 geographically defined PSUs for the entire country; subsequently, 20 final target households were sampled for each PSU as secondary sampling units (ID_fam) [16]. The sample weights were calculated using the inverse of selection probabilities and the inverse of response rates that were adjusted to gender- and age-specific Korean populations (poststratification). Therefore, the sampled individuals accurately represented the Korean population.

Depression Case Definition

Positive depression diagnosed by a physician (DF2_dg) and positive present depression (DF2_pr) represented the positive case “current depression,” and these indicated that the individual had a diagnosis of depression and was currently experiencing depression, respectively. In terms of the negative case, both DF2_dg and DF2_pr were negative and represented “no lifetime depression,” which indicated that the individual had never experienced depression. DF2_dg and DF2_pr were reported by the participants in this survey.

XGBoost-Derived Model for Depression

We included as many variables as possible (Textbox 1) and imputed missing values using univariate imputation (filling with the last valid value for categorical or discrete variables) or multivariate imputation for missing continuous variables. The nutrient variables were discretized into three bins as follows: low, medium, and high. We converted categorical variables to numerical variables using one-hot encoding and dropped reference categories. We standardized variables, except for a binary dummy variable, by removing the mean and scaling them to unit variance before machine learning.

Textbox 1. Study variables of the Korea National Health and Nutrition Examination Survey data.

Health Examination Data

Physical examination

- Waist circumference, BMI, regularity of pulse, and periodontal disease

Blood tests

- Anemia (hemoglobin <13.0 g/dL for men; hemoglobin <12.0 g/dL for nonpregnant women; hemoglobin <11.0 g/dL for pregnant women), white blood cell count, platelet count, red blood cell count, aspartate aminotransferase, alanine aminotransferase, creatinine, urea nitrogen, hepatitis B surface antigen, and hepatitis C antibody

Fasting (≥8 hours) blood tests

- Sugar level, high-density lipoprotein cholesterol (HDL), and triglyceride (TG)

Urine strip test

- pH, nitrite, specific gravity, protein, glucose, ketones, bilirubin, blood, urobilinogen

Pulmonary function test

Hypercholesterolemia: total cholesterol (TC) ≥240 mg/dL or lipid-lowering medication

Hypertension: systolic pressure ≥140 mmHg or diastolic pressure ≥90 mmHg or medication

Diabetes mellitus: fasting blood sugar level ≥126 mg/dL, diagnosis, medications, or insulin injections

Prediabetes: fasting blood sugar level ≥100 and <126 mg/dL

Low-density lipoprotein cholesterol (LDL) = TC – (HDL + TG/5), where TG ≤400 mg/dL

Health Survey Data

- Age, marital status, educational level, occupational class, household income, weight changes in the past year, alcohol consumption frequency, sleep hours per day, perceived stress, current smoking status, walking for more than 10 minutes (days/week), muscle-strengthening activities (days/week), childbirth experience, and EuroQol five-dimension three-level (EQ-5D-3L) questionnaire (mobility, usual activities, self-care, and pain/discomfort; three levels for each) [17]

Morbidity

- Chronic obstructive pulmonary disease, stroke, ischemic heart disease, osteoarthritis, rheumatoid arthritis, pulmonary tuberculosis, asthma, thyroid disease, gastric cancer, hepatoma, colon cancer, breast cancer, cervix cancer, lung cancer, thyroid cancer, atopic dermatitis, renal failure, hepatitis B, hepatitis C, and liver cirrhosis

Oral health

- Difficulty chewing, caries treatment within the last 1 year, root canal treatment within the last 1 year, and prosthetic treatment within the recent 1 year

Food Frequency Survey Data (daily intake)

- Carbohydrates, proteins, saturated fatty acids, monounsaturated fatty acids, polyunsaturated fatty acids, n3 fatty acids, n6 fatty acids, n6:n3 fatty acid ratio, cholesterol, fiber, vitamin A, vitamin B1, vitamin B2, vitamin C, niacin, iron, calcium, potassium, phosphorus, and sodium

The training set included 80% “current depression” and 80% “no lifetime depression” (10,076 cases). The remaining cases were assigned to the test set (2520 cases). We used TPOT to choose the best XGBoost hyperparameters from the training set to predict “current depression” from “no lifetime depression.” Then, the model features were selected further using the feature importance of XGBoost. The XGBoost model was explained using shapley additive explanation (SHAP) values that showed each feature’s impact on the model prediction [18].

The weighted area under the receiver operating characteristic curve (AUC) of the final model was calculated using the sample weight [19]. The optimal threshold was computed as the prediction probability at which Youden’s index (sensitivity + specificity – 1) was the maximum on the test set. We also estimated the weighted AUC of the model for the Patient Health Questionnaire-9 (PHQ-9) score (depression when score ≥10) instead of the reported depression.

Statistical Test for Model Factors and Exploring Nonmodel Factors

Survey-weighted multiple logistic regression was performed using the model features from all samples (training and test sets). We defined the “direct” factor of depression as the model feature whose multiple logistic regression coefficient significantly differs from zero ($P<.05$); the other model features were designated as “indirect” factors. Subsequently, we performed a statistical test on factors excluded by the model to identify significant “nonmodel” factors that were not chosen by XGBoost as optimal for predicting depression. Nonmodel factors whose coefficients of multiple logistic regression with the model features were significantly different from zero (P<.05 with Bonferroni correction) were chosen.

Network Analysis for Depression-Related Factors

A correlation matrix of model features and nonmodel factors was generated using all samples. Based on the correlation matrix, a thresholded EBICglasso (extended Bayesian information criterium graphical lasso) network was plotted, in which the network model was estimated using graphical LASSO regularization with EBIC model selection. Each edge weight was the correlation coefficient between two nodes after controlling all other network correlations [20,21].

Three centrality indices (strength, closeness, and betweenness) were computed. Strength centrality is the absolute sum of the edge weights connected to the node. Closeness centrality is the sum of the shortest distances from the node to all other network nodes. Betweenness centrality is the number of times when the node lies on the shortest path between two other nodes [21,22].

Confounding or Interaction Effects on Indirect Risk Factors

If an indirect risk factor was positively connected to direct risk factors, the direct factor confounding effects were tested. If the indirect factor's coefficient on survey-weighted multiple logistic regression became significant without the direct factors, we assumed they were the indirect factor's confounders.

If an indirect risk factor was positively connected to a direct preventive factor or negatively related to a direct risk factor, the interaction effect was tested using the interaction term on survey-weighted multiple logistic regression.

Statistics and Software

We used the following two programming languages: Python (version 3.8.5, Python Software Foundation) and R (version 3.6.3, R Foundation for Statistical Computing).

Scikit-Learn Tools

The following scikit-learn tools [23] were applied: for imputing missing continuous values, *IterativeImputer* class with the ExtraTreesRegressor estimator; for discretization, *KBinsDiscretizer* with a k-means strategy based on a k-means clustering procedure; for standardization, *StandardScaler*; and for weighted AUC, sensitivity, specificity, and threshold, *roc_auc_score* and *roc_curve*.

TPOT

TPOT starts from a collection of random models (first generation). Subsequently, those with higher performance are chosen and copied into the next generation's population. The offspring crossover with other offspring or are randomly changed by mutation. The algorithm repeats this evaluate-select-crossover-mutate process for multiple generations. Finally, the best model is selected from the run [13].

We used TPOT to tune the hyperparameters of the XGBClassifier, such as the number of boosting rounds ("n_estimators"), boosting learning rate ("learning_rate"), maximum tree depth for base learners ("max_depth"), minimum sum of instance weight needed in a child ("min_child_weight"), minimum loss reduction required to make a further partition on a leaf node of the tree ("gamma"), subsample ratio of the training instance ("subsample"), subsample ratio of columns when constructing each tree ("colsample_bytree"), L1 regularization term on weights ("reg_alpha"), and L2 regularization term on weights ("reg_lambda"). To control the balance of positive and negative weights, we set "scale_pos_weight" as the number of negative cases/the number of positive cases. The "objective" option was set to a default value, "binary:logistic" (ie, logistic regression for binary classification).

Two parameters of TPOTClassifier, namely, number of iterations to run the XGBoost optimization process ("generation") and number of individuals to retain in the genetic programming population every generation ("puplation_size"), were set to 100 (default) each. Finally, TPOT evaluated models using AUC ("roc_auc").

Extreme Gradient Boosting (XGBoost) and Feature Selection

XGBoost incorporates an ensemble technique (boosting) of adding new models to correct errors made by previous models. When adding new models, gradient boosting uses a gradient descent algorithm to minimize the loss. More details on XGBoost's working principles can be found in published articles [24]. For feature selection, we sorted the model features according to their importance (importance_type = "gain," the average gain of splits that use the feature) and selected a subset of features by omitting the least important feature. If the AUC of the model with the subset of features was not compromised or became higher, we omitted the next least important feature. Subsequently, we retested it until the model performance reached the best AUC.

R Packages

The *Survey* package allows to specify a multistage sampling survey design and provides functions to estimate total population counts, means, and variances for a survey sample [25]. In this study, the survey design of the KNHANES was applied using the variables *psu* for PSU, *ID_fam* (family ID) for the secondary sampling unit, *kstrata* for strata, and sample weight (*wt_tot*/5) for weight. Additionally, we used the "svyglm" function from the *survey* package for survey-weighted multiple logistic regression. The odds ratio was calculated by exponentiating the coefficient of multiple logistic regression.

Receiver operating characteristic (ROC) curves were generated using the *ggplot2* package [26], and the weighted AUC was calculated by using the "WeightedAUC" function with the sample weight (*WeightedROC* package) [19].

Network Graph

For a correlation matrix in network analysis, a correlation between two variables was calculated using the wtd.cor function with the sample weight for "weight" in R's weights package, which produced weighted Pearson correlation coefficients for survey data [27]. The network graph was generated using the qgraph function in the *qgraph* package of R by setting the graph argument to “glasso” and layout to “spring” with threshold [20]. For node sizes in the network, the effective size of the variable’s odds ratio was calculated by exponentiating the absolute coefficient of multiple logistic regression. For the calculation of centrality indices, the qgraph object and the centralityTable function in the *qgraph* package of R were used [20].

Results

XGBoost-Derived Model and the Performance

The population was estimated to be 22,262,880 (SE 355,977) (97.3%) for no depression reported during lifetime and 616,082 (SE 40,273) (2.7%) for current depression (complex-survey-design-based estimation). After TPOT training, the number of XGBoost model features was reduced from 120 to 81 (33% reduction) by the L1 regularization term of XGBoost. Furthermore, we decreased the feature number to 18 (78% reduction) by applying feature selection without performance loss.

The final model features included mental health, quality of life (QoL), socioeconomic factors, morbidity, sex, marital status, urinalysis, and health behavior (Figure 2). Nutrition factors failed to be included in the model. For the test set, the weighted AUC of the model was 0.86 for reported depression and 0.82 for PHQ-9 score depression (Figure 2). The model’s accuracy was 0.82 at the best threshold of the test data (Table 1).

Figure 2. XGBoost model for the depression and performance test. Red dots represent positive for a binary factor or high feature values for a continuous factor on the beeswarm plot of shapley additive explanation (SHAP) values (log odds of the current depression) for the training data (A). On the weighted receiver operating characteristic curve (ROC) (B), sensitivity is 0.78 and specificity is 0.82 at the best threshold of probability of 0.461 for test samples. The model is also tested using the Patient Health Questionnaire-9 (depression when score ≥10) for 6098 samples (C). AUC: area under the curve.

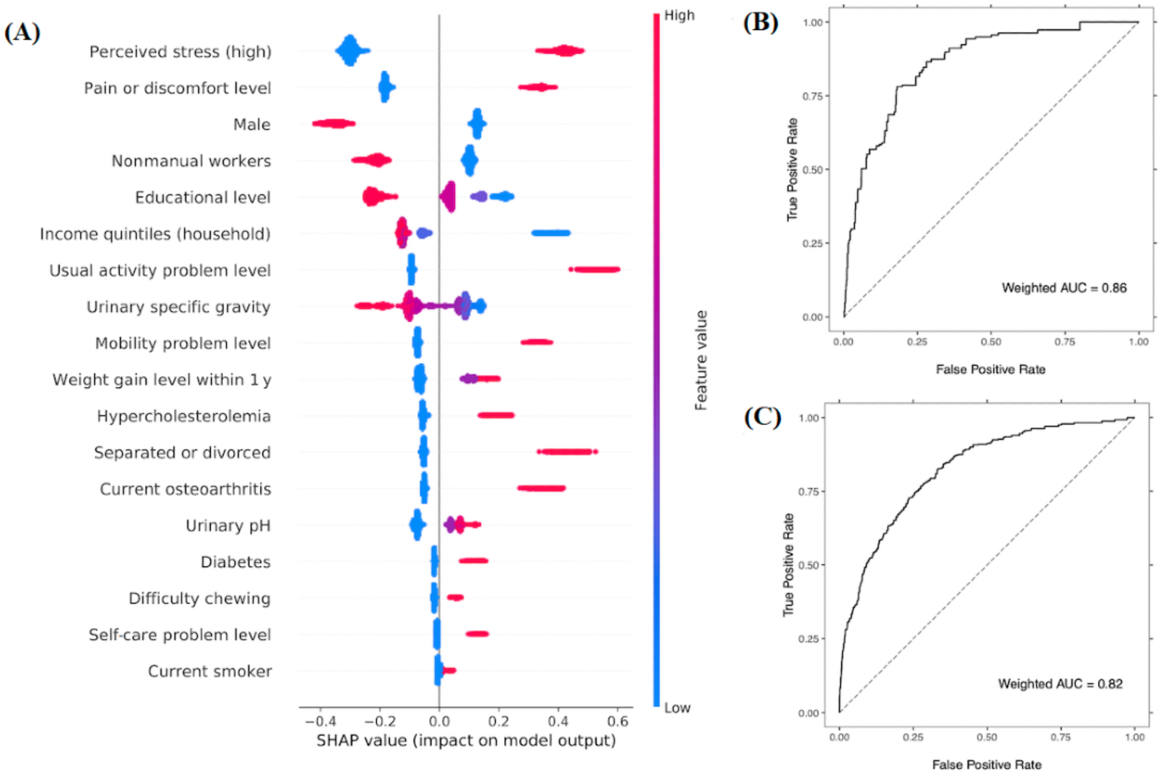


Table 1. Confusion matrix for the test data set.

Actual depression	Predicted depression[a]	
	Current	No lifetime
Current	103,182	28,933
No lifetime	804,195	3,657,604

[a] Case numbers are estimated using complex-survey-design weights at the best threshold.

Statistical Significance of Depression-Associated Factors

Perceived stress and current asthma were top-ranked statistically significant risk factors (Table 2). Triglycerides, current asthma, and farm workers were nonmodel factors having a significant association with depression and added on the network with model factors. Eight features out of 18 model factors were not statistically significant and classified as indirect (Table 2).

Table 2. Survey-weighted multiple logistic regression analysis of depression-related factors.

Variable	No lifetime depression[a] (N=22,262,880)	Current depression[a] (N=616,082)	Odds ratio (95% CI)[b]	*P* value
High perceived stress (≥ 3/4 points), n (%)				
Yes	5,999,970 (26.95%)	372,138 (60.40%)	3.3 (2.5-4.3)	<.001
No (reference)	16,262,910 (73.05%)	243,944 (39.60%)	N/A[c]	N/A
Sex/gender, n (%)				
Male	11,072,884 (49.74%)	165,770 (26.91%)	0.5 (0.3-0.7)	<.001
Female (reference)	11,189,996 (50.26%)	450,312 (73.09%)	N/A	N/A
Marital status, n (%)				
Separated or divorced	852,170 (3.83%)	105,567 (17.14%)	2.2 (1.4-3.5)	<.001
Single	6,153,788 (27.64%)	133,952 (21.74%)	1.4 (0.9-2.2)	.11
Widowed	397,387 (1.78%)	36,923 (5.99%)	1.4 (0.8-2.6)	.25
Married (reference)	14,859,535 (66.75%)	339,640 (55.13%)	N/A	N/A
Occupation, n (%)				
Nonmanual	10,304,597 (46.29%)	161,758 (26.26%)	0.6 (0.4-0.8)	<.001
Manual	4,394,739 (19.74%)	102,028 (16.56%)	0.7 (0.4-1.0)	.06
Farm[d]	498,239 (2.24%)	10,848 (1.76%)	0.4 (0.2-0.8)	.02
Unemployed (reference)	7,065,305 (31.73%)	341,448 (55.42%)	N/A	N/A
Asthma[d], n (%)				
Current	205,771 (0.92%)	34,769 (5.64%)	3.1 (1.5-6.5)	.002
Past	303,213 (1.36%)	10,387 (1.69%)	1.3 (0.5-3.2)	.57
No lifetime (reference)	21,753,896 (97.72%)	570,926 (92.67%)	N/A	N/A
Hypercholesterolemia, n (%)				
Yes	2,963,312 (13.31%)	146,742 (23.82%)	1.4 (0.9-1.9)	.10
No (reference)	19,299,568 (86.69%)	469,340 (76.18%)	N/A	N/A
Osteoarthritis, n (%)				
Current	825,130 (3.71%)	100,307 (16.28%)	1.4 (0.9-2.1)	.17
Past	155,266 (0.70%)	8,385 (1.36%)	0.7 (0.1-4.6)	.71
No lifetime (reference)	21,282,484 (95.59%)	507,390 (82.36%)	N/A	N/A
Diabetes, n (%)				
Yes	1,466,133 (6.59%)	84,663 (13.74%)	1.4 (0.9-2.2)	.11
Prediabetes	4,504,230 (20.23%)	125,783 (20.42%)	1.0 (0.7-1.4)	.81
No (reference)	16,292,517 (73.18%)	405,636 (65.84%)	N/A	N/A
Difficulty chewing, n (%)				
Yes	3,398,898 (15.27%)	173,323 (28.13%)	0.9 (0.7-1.3)	.65
No (reference)	18,863,982 (84.73%)	442,759 (71.87%)	N/A	N/A
Current smoker, n (%)				
Yes	5,178,518 (23.26%)	139,892 (22.71%)	1.2 (0.8-1.8)	.47
No	17,084,362 (76.74%)	476,190 (77.29%)	N/A	N/A
Pain or discomfort level (1 to 3 points), mean (SE)	1.2 (0.004)	1.6 (0.034)	1.3 (1.1-1.5)	<.001
Fasting serum triglyceride (mg/dL)[d], mean (SE)	134.8 (1.3)	161.8 (12.0)	1.2 (1.1-1.3)	.002
Weight gain level within 1 year (0 to 3 points), mean (SE)	0.4 (0.009)	0.6 (0.047)	1.2 (1.1-1.3)	.003

Variable	No lifetime depression[a] (N=22,262,880)	Current depression[a] (N=616,082)	Odds ratio (95% CI)[b]	*P* value
Urine specific gravity, mean (SE)	1.020 (0.00007)	1.018 (0.0004)	0.8 (0.7-0.9)	.003
Usual activity problem level (1 to 3 points), mean (SE)	1.0 (0.002)	1.3 (0.029)	1.1 (1.0-1.3)	.006
Income quintiles (household) (1 to 5 points), mean (SE)	3.4 (0.02)	2.6 (0.09)	0.8 (0.7-0.9)	.007
Educational level (1 to 4 points), mean (SE)	3.2 (0.01)	2.6 (0.06)	0.8 (0.7-1.0)	.01
Urine pH, mean (SE)	5.7 (0.009)	5.8 (0.050)	1.1(1.0-1.3)	.07
Mobility problem level (1 to 3 points), mean (SE)	1.1 (0.002)	1.3 (0.032)	1.1 (1.0-1.2)	.22
Self-care problem level (1 to 3 points), mean (SE)	1.0 (0.001)	1.1 (0.017)	1.0 (0.9-1.1)	.98
Age (years)[e], mean (SE)	40.7 (0.2)	45.1 (0.8)	1.1 (0.8-1.3)	.61

[a] Population counts (n), means, and standard errors are estimated using complex-survey-design weights.

[b] For continuous factors, the value was standardized by removing the mean and scaling to unit variance.

[c] N/A: not applicable.

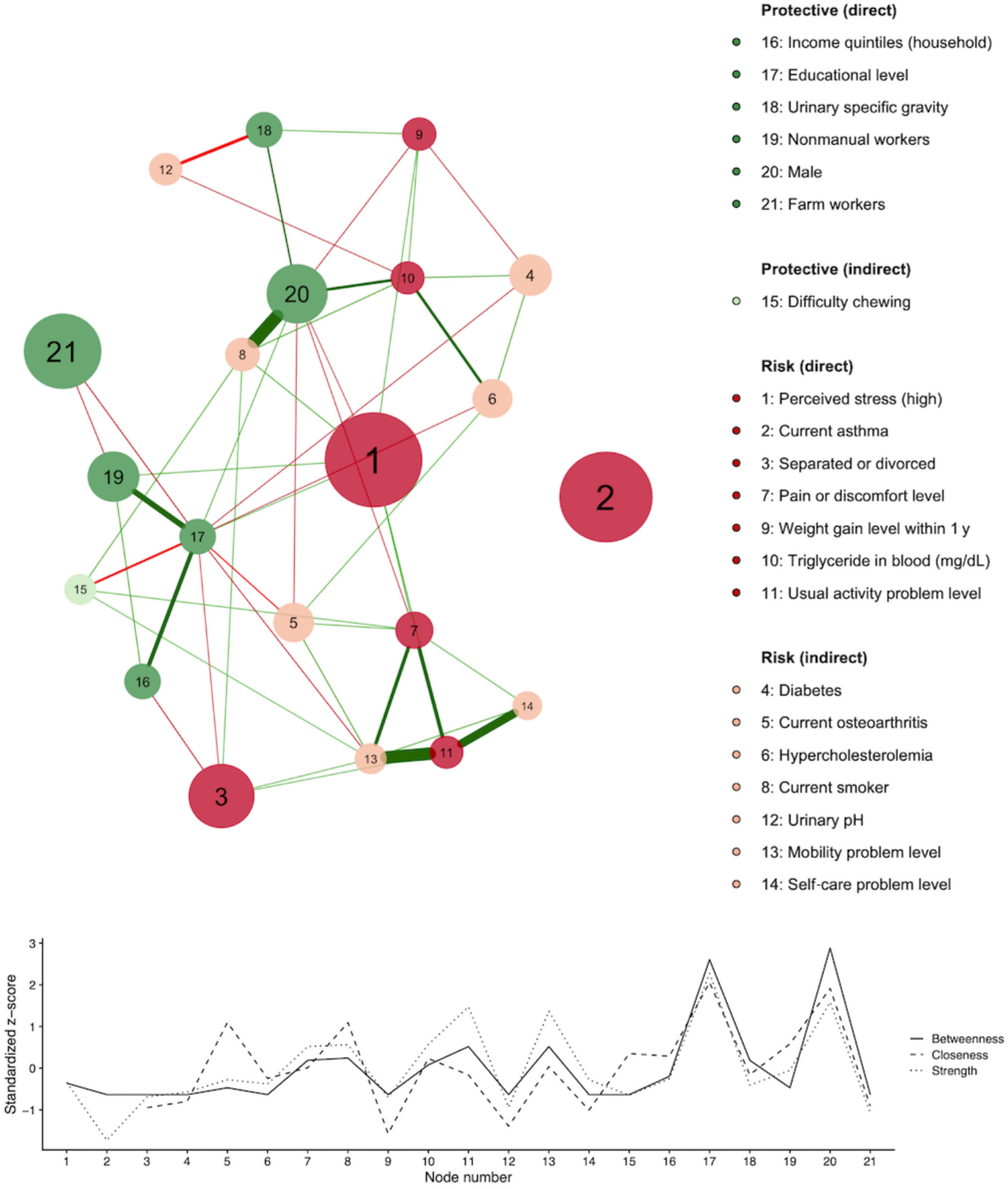
[d] Nonmodel factors were found by controlling for model features and age.

[e] Not a depression-related factor, but included to control confounding effects of age.

High-Centrality Nodes and Clusters in the Depression Network

Based on the centrality indexes, the network had two high-centrality nodes, namely, “educational level” and “male,” which meant both nodes were highly connected with various factors (Figure 3). Some factors were strongly connected to form a cluster (eg, the socioeconomic status [SES] cluster consisted of nonmanual workers, farm workers, educational level, and household income) (Figure 3). The QoL cluster comprised four EuroQol five-dimension three-level (EQ-5D-3L) domains (“pain or discomfort level,” “usual activity problem level,” “mobility problem level,” and “self-care problem level”) (Figure 3).

Figure 3. Partial correlation network graph and centrality indices for depression-associated factors. The factors can be positively (risk) or negatively (protective) related to current depression. Node size is proportional to the effective size of the odds ratio. Green and red edges represent positive and negative correlations, respectively. The edge with the highest absolute weight has full-color saturation and the widest width.



Confounders and Interaction Effects of Indirect Factors

Two indirect factors, namely, hypercholesterolemia and diabetes, were positively connected with triglyceride on the network, and triglyceride was a confounder of hypercholesterolemia and diabetes (Table 3). Another indirect factor, current osteoarthritis, became statistically significant after excluding its positive associates, namely, “pain or discomfort level” and “mobility problem level” (Table 3).

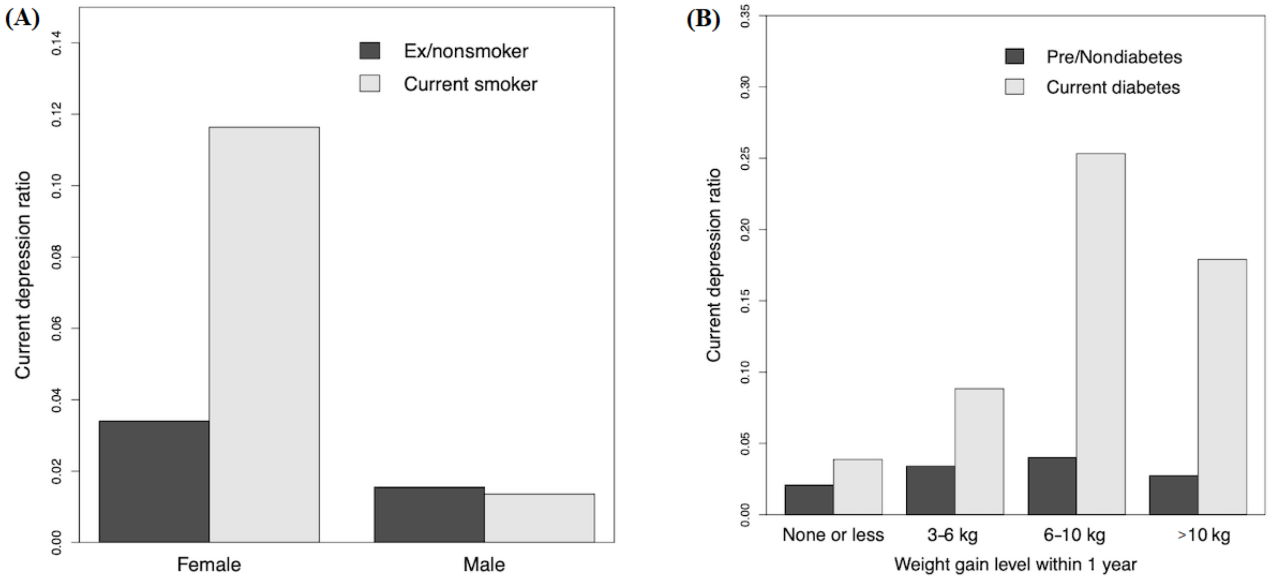
Table 3. Possible confounders on indirect factors in depression.

Indirect factors	Confounders[a]	Odds ratio (95% CI)	
		With confounders	Without confounders
Current osteoarthritis	Pain or discomfort level and mobility problem level	1.4 (0.9-2.1)	1.6 (1.1-2.5)
Hypercholesterolemia	Triglyceride	1.4 (1.0-1.9)	1.5 (1.0-2.1)
Diabetes	Triglyceride and hypercholesterolemia	1.4 (0.9-2.2)	1.7 (1.1-2.6)

[a] A group of confounders to make a coefficient of the indirect factor statistically insignificant.

Another indirect factor, “current smoker,” interacted with sex and was statistically significant in females (odds ratio 2.2, 95% CI 1.3-3.7; *P*=.003) (Figure 4). Current diabetes interacted with the weight gain level that was not statistically significant for prediabetes or nondiabetes (*P*=.05) (Figure 4).

Figure 4. The proportion of current depression according to sex and smoking (A) and according to weight gain and diabetes (B). The interaction effects between them are significant (*P*=.001 and .01, respectively; survey-weighted logistic regression).



Urine Specific Gravity and Depression

We found an interesting relationship between urine specific gravity (USG) and depression. Because USG was associated with the “male” node on the network (Figure 3), we stratified it into males and females and plotted it according to age. USG was lower in “current depression” than in “no lifetime depression,” especially in females who were in their early 50s or younger (Figure 5). However, the daily water intake was not higher in “current depression” than in “no lifetime depression” (Figure 6).

Figure 5. Linear regression plots between age and urine specific gravity in females and males. In the images, 95% confidence intervals for the regression estimate are drawn using translucent bands around the regression line.

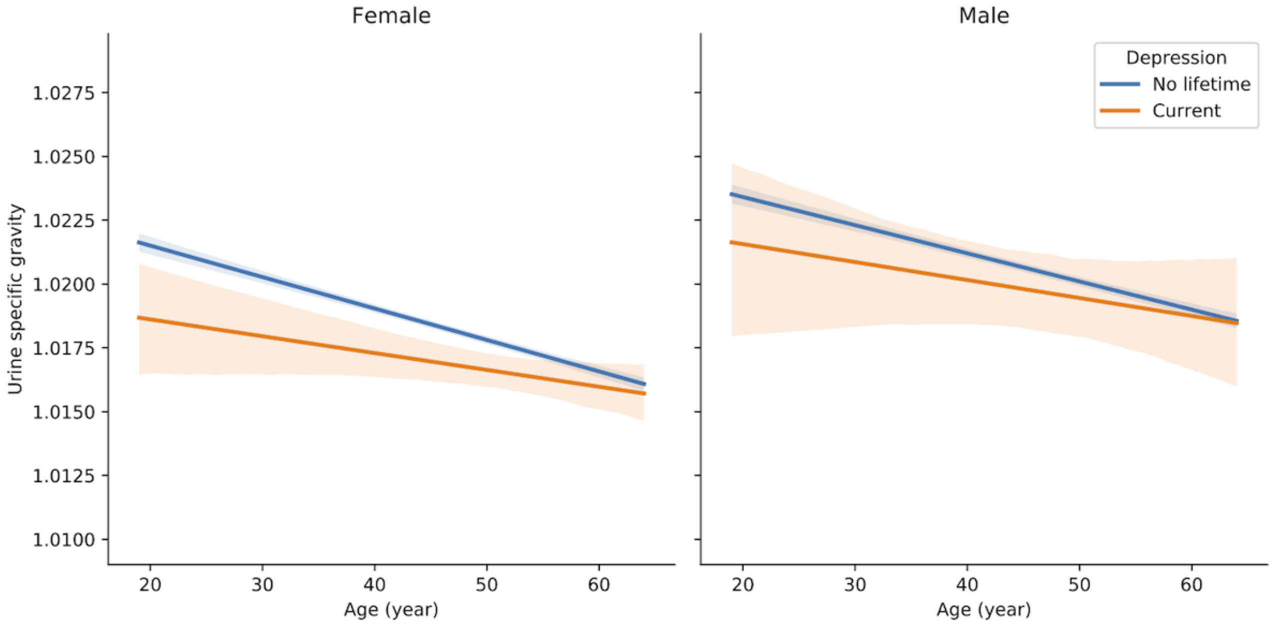
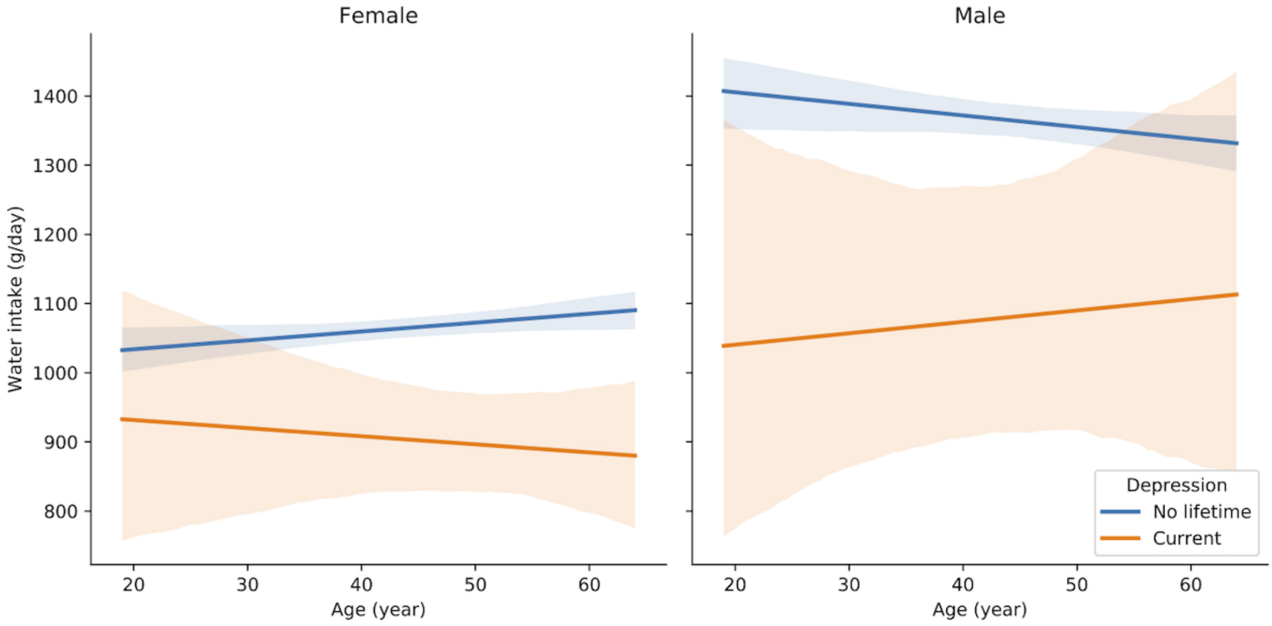


Figure 6. Linear regression plots between age and daily water intake in females and males. In the images, 95% confidence intervals for the regression estimate are drawn using translucent bands around the regression line.



Discussion

Addressing Class Imbalance and Generalization Problems

Our data set was highly imbalanced, with a 2.7% prevalence of depression. Therefore, it may be difficult for a machine learning algorithm to predict depression because this diagnosis is not sufficiently represented [28]. We used the “scale_pos_weight” parameter of XGBoost that enables class-weighted or cost-sensitive learning to make the model more sensitive to misclassification in the minority class [12]. The performance of our XGBoost model showed that the TPOT-trained class-weighted XGBoost model was working well for our imbalanced classification.

There would have been a generalization problem with this survey data because we did not train the model using the sample weight. However, the weighted AUC for the test set (0.86) was not lower than the mean unweighted AUC for the training set (0.84, SD 0.02; seven repeats of five-fold cross-validation). Therefore, our model was not overfitted, and we expect that our model can be applied to the general population.

Reliability of Self-Reported Depression

A model may be less reliable when it is based on self-reported depression than on clinically diagnosed depression. However, for PHQ-9 scores, a clinical module, the model performance was still satisfactory when the model was retested with available cases (Figure 2). PHQ-9 is a validated screening tool for depression and is known to exhibit an acceptable diagnostic accuracy for cutoff scores between 8 and 11 [29,30]. It might be possible because we strictly defined depression by requiring both conditions (ie, current symptoms and a physician's diagnosis). In addition, we restricted "negative depression" only to the cases when depression was never diagnosed, and respondents with a past diagnosis of depression were excluded.

Combination of the XGBoost Model and Statistical Analysis: Feature Reduction, Indirect Factors, and Nonmodel Factors

XGBoost L1 regularization did not ultimately reduce the feature number; however, feature selection using XGBoost feature importance could decrease the feature number more at a higher reduction rate. Therefore, we suggest feature reduction using XGBoost feature importance to obtain a sparser model that contains the most principle factors of the disease.

Because of the relatively small number of predictors, a sparse model is interpretable without irrelevant features, which could be shown by the impact on the depression of each feature (Figure 2). Furthermore, we performed statistical tests on the sparse model features using multiple logistic regression at a lower risk of multicollinearity or redundancy problems. Overall, 44% of the model features were not statistically significant, and we classified them as indirect factors because they affected depression without a direct statistical association with depression.

Additionally, nonmodel factors (triglyceride, current asthma, and farm worker) exhibited statistical significance by multiple regression with depression model factors. By controlling principle depression factor effects (eg, confounding effects), nonmodel factors' significance was reliable. Therefore, we might use a sparse disease model as a testing tool for candidate factors.

Elements of the Depression Network: Node Size and Edge

Our network was composed of differently sized nodes (statistical strength of the association) and edges (net correlation between two factors). "Perceived stress" and "current asthma" were prioritized in depression risk factor control because of their large nodes (Figure 3). In addition to genetic factors, stressful events affect the onset of depression and cause depression through psychological stress responses, such as activating the hypothalamic-pituitary-adrenal axis [31]. Additionally, the network indicated that "perceived stress" could accompany other risk factors, such as "pain and discomfort" and "weight gain," and we might consider them together for risk factor control (Figure 3). In contrast, current asthma was an independent node, which suggests asthma may directly link with depression (eg, asthma medications or dysregulation of specific stress-sensitive biological processes) and be an individual control target [32].

Depression Network Centrality and Clusters: Indication of Predisposing Factors and Factor Groups

In terms of the network's centrality indexes, "male" and "educational level" showed the highest values. Scale-free networks are characterized by growth and preferential attachment in which earlier nodes in the network increase their connectivity at a higher rate [33]. Therefore, high-centrality nodes might be preceding factors. For example, the high centrality of "educational level" might explain that a higher educational level has accumulated protective effect throughout life [34]. Therefore, the network centrality suggested that gender and the education level would be predisposed to biological and socioeconomic depression factors. Statistically, gender and the education level might be must-have covariates because of their multiple relations to other factors.

Clusters in the disease network can be potent risk intervention targets because factors in a cluster are connected and controlled together [35]. The SES cluster on our depression network is reported as the most prominent risk among psychosocial risks [35], and low SES is common in depression because of poor care, low treatment compliance, and job strain [36]. For the QoL cluster, a study reported that the usual activity problem and the pain level of EQ-5D are mainly affected by depression, and the treatment of depression improves the EQ-5D index score [37]. Depression symptoms were negatively connected to health-related QoL in the network in a previous study [38].

Confounding and Interaction Effects of Indirect Risk Factors

If an indirect risk factor is connected to a direct risk factor, the direct factor could be the confounder. Furthermore, if the direct factor is on the causal pathway, it can be a mediator of the indirect factor action. We found potential confounders of current osteoarthritis, hypercholesterolemia, and diabetes using the network (Table 3). Osteoarthritis causes pain and physical disability, which can reduce QoL and lead to the development of depression [39]. This can be seen in our network based on the relationships among current osteoarthritis, pain, and mobility problems in the QoL cluster (Figure 3).

Obesity and metabolic syndrome may mediate the relationship between diabetes and depression [40]. This fact could be captured in our network because diabetes was associated with weight gain, higher triglycerides, and hypercholesterolemia (Figure 3). Additionally, the elevated triglyceride level was a confounder of hypercholesterolemia (Table 3), which may explain a recent meta-analysis reporting that the first episode of major depressive disorder is associated with elevated triglyceride levels, not low-density lipoprotein cholesterol or total cholesterol levels [41]. Previous studies reported that an elevated triglyceride level is associated with depression and suicidality in men [3] or women [42]. However, there was no interaction effect between "male" and triglyceride level in our study (*P*=.82).

If an indirect risk factor is connected to a direct protective factor or negatively associated with a direct risk factor, its role could be revealed by the interaction effect. "Current smoker," which is an indirect factor strongly connected to "male," was a significant risk factor in females, but not in males (Figures 3 and 4). Smoking cessation might be considered, especially for female depression patients [43]. Weight gain occurred in depression in current diabetes, but not clearly in prediabetes or nondiabetes (Figure 4).

USG as a Novel Depression-Related Factor

To the best of our knowledge, USG was a novel factor of depression, and the level was low in female depressive subjects (Figure 5). USG correlates with urine osmolality and reflects the subject's hydration status or the kidney's concentrating ability [44]. However, our study showed that low USG in depression might not be caused by hydration because daily water intake was not low in depression (Figure 6). One study suggested that patients with depression concentrate the urine less well and excrete less solute in the urine [45].

Limitations

This cross-sectional study involved people aged 19 to 64 years and could analyze only associations, not causalities. Therefore, the associations established in this study, such as weight gain in diabetes, elevated triglyceride, low USG, and smoking in females, require future clinical research to prove their efficacy in depression control. We used Bonferroni correction for nonmodel factors, and this conservative method might miss other possibly significant factors. Several cancers (hepatoma, gastric cancer, colon cancer, breast cancer, and lung cancer), renal failure, pulmonary tuberculosis, liver cirrhosis, and hepatitis C might not have enough positive cases to be tested. Finally, psychiatrists' structured interviews to diagnose depression would be more valid than the self-reported approach to identify depression used in this study.

Conclusions

We successfully created a sparse model for depression using TPOT-assisted XGBoost training and feature selection based on the feature importance of XGBoost from a large number of variables in KNHANES data. Because of the data-driven approach, we could discover a novel factor. We constructed a network of the depression-associated factors using association strength and interfactor correlations. The model factors were classified into direct and indirect according to their statistical significance, and the role of indirect factors was explained by confounding or interaction effects. The network also indicated predisposing factors by high centrality and cluster factors by a closely connected edge. Therefore, XGBoost and network analysis can be useful for discovering and understanding disease-associated factors in epidemiological studies.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2019R1C1C1007663, 2020R1F1A106842312, and 2019R1A2C1084611).

Authors' Contributions

HWH and JIK contributed equally as corresponding authors.

Conflicts of Interest

None declared.

References

1. National Collaborating Centre for Mental Health. Depression: The NICE Guideline on the Treatment and Management of Depression in Adults (Updated Edition) (Nice Guidelines). Leicester, UK: British Psychological Society; 2010.
2. Hasler G. Pathophysiology of depression: do we have any solid evidence of interest to clinicians? World Psychiatry 2010 Oct;9(3):155-161 [FREE Full text] [doi: 10.1002/j.2051-5545.2010.tb00298.x] [Medline: 20975857]
3. Kim B, Park EY. The combined effect of socioeconomic status and metabolic syndrome on depression: the Korean National Health and Nutrition Examination Survey (KNHANES). BMC Public Health 2020 May 04;20(1):617 [FREE Full text] [doi: 10.1186/s12889-020-08778-3] [Medline: 32366283]
4. Jager KJ, Zoccali C, Macleod A, Dekker FW. Confounding: what it is and how to deal with it. Kidney Int 2008 Feb;73(3):256-260 [FREE Full text] [doi: 10.1038/sj.ki.5002650] [Medline: 17978811]
5. Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. Am J Epidemiol 2008 Mar 01;167(5):523-9; discussion 530. [doi: 10.1093/aje/kwm355] [Medline: 18227100]
6. Stoltzfus JC. Logistic regression: a brief primer. Acad Emerg Med 2011 Oct;18(10):1099-1104 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01185.x] [Medline: 21996075]
7. Lima E, Davies P, Kaler J, Lovatt F, Green M. Variable selection for inferential models with relatively high-dimensional data: Between method heterogeneity and covariate stability as adjuncts to robust selection. Sci Rep 2020 May 14;10(1):8002 [FREE Full text] [doi: 10.1038/s41598-020-64829-0] [Medline: 32409668]
8. Dankers FJWM, Traverso A, Wee L, van Kuijk SMJ. Prediction Modeling Methodology. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham: Springer; 2019:101-120.
9. Tlachac ML, Rundensteiner EA. Depression Screening from Text Message Reply Latency. Annu Int Conf IEEE Eng Med Biol Soc 2020 Jul;2020:5490-5493. [doi: 10.1109/EMBC44109.2020.9175690] [Medline: 33019222]

10. Wang S, Pathak J, Zhang Y. Using Electronic Health Records and Machine Learning to Predict Postpartum Depression. Stud Health Technol Inform 2019 Aug 21;264:888-892. [doi: 10.3233/SHTI190351] [Medline: 31438052]
11. Gu S, Zhou J, Yuan C, Ye Q. Personalized prediction of depression in patients with newly diagnosed Parkinson's disease: A prospective cohort study. J Affect Disord 2020 May 01;268:118-126. [doi: 10.1016/j.jad.2020.02.046] [Medline: 32158001]
12. Brownlee J. XGBoost With Python: Gradient Boosted Trees with XGBoost and scikit-learn. San Juan, PR: Machine Learning Mastery; 2018.
13. Olson RS, Moore JH. TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning. In: Hutter F, Kotthoff L, Vanschoren J, editors. Automated Machine Learning. The Springer Series on Challenges in Machine Learning. Cham: Springer; 2019:151-160.
14. Kweon S, Kim Y, Jang M, Kim Y, Kim K, Choi S, et al. Data resource profile: the Korea National Health and Nutrition Examination Survey (KNHANES). Int J Epidemiol 2014 Feb;43(1):69-77 [FREE Full text] [doi: 10.1093/ije/dyt228] [Medline: 24585853]
15. Yoon KC, Choi W, Lee HS, Kim S, Kim S, Kim CY, et al. An Overview of Ophthalmologic Survey Methodology in the 2008-2015 Korean National Health and Nutrition Examination Surveys. Korean J Ophthalmol 2015 Dec;29(6):359-367 [FREE Full text] [doi: 10.3341/kjo.2015.29.6.359] [Medline: 26635451]
16. Korea National Health and Nutrition Examination Survey. Korea Centers for Disease Control and Prevention. URL: https://knhanes.kdca.go.kr/knhanes/eng/index.do [accessed 2021-06-05]
17. EuroQol. EQ-5D-3L. URL: https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/ [accessed 2021-06-05]
18. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
19. WeightedROC: Fast, Weighted ROC Curves. r-project. URL: https://CRAN.R-project.org/package=WeightedROC [accessed 2021-06-05]
20. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: Network Visualizations of Relationships in Psychometric Data. J. Stat. Soft 2012 Sep 1;48(4):1-18. [doi: 10.18637/jss.v048.i04]
21. Pereira-Morales AJ, Adan A, Forero DA. Network analysis of multiple risk factors for mental health in young Colombian adults. J Ment Health 2019 Apr;28(2):153-160. [doi: 10.1080/09638237.2017.1417568] [Medline: 29265896]
22. Robinaugh DJ, Millner AJ, McNally RJ. Identifying highly influential nodes in the complicated grief network. J Abnorm Psychol 2016 Aug;125(6):747-757 [FREE Full text] [doi: 10.1037/abn0000181] [Medline: 27505622]
23. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12:2825-2830 [FREE Full text]
24. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August, 13, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
25. Lumley T. Complex Surveys: A Guide to Analysis Using R. Hoboken, NJ: John Wiley & Sons, Inc; 2010.
26. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Cham: Springer; 2016.
27. Pasek J. weights: Weighting and Weighted Statistics. r-project. URL: https://CRAN.R-project.org/package=weights [accessed 2021-06-05]
28. Brownlee J. Imbalanced Classification with Python. San Juan, PR: Machine Learning Mastery; 2020.
29. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
30. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 2012 Feb 21;184(3):E191-E196 [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]
31. Yang L, Zhao Y, Wang Y, Liu L, Zhang X, Li B, et al. The Effects of Psychological Stress on Depression. Curr Neuropharmacol 2015;13(4):494-504 [FREE Full text] [doi: 10.2174/1570159x1304150831150507] [Medline: 26412069]
32. Van Lieshout RJ, Bienenstock J, MacQueen GM. A review of candidate pathways underlying the association between asthma and major depressive disorder. Psychosom Med 2009 Feb;71(2):187-195. [doi: 10.1097/PSY.0b013e3181907012] [Medline: 19073754]
33. Barabasi, Albert R. Emergence of scaling in random networks. Science 1999 Oct 15;286(5439):509-512 [FREE Full text] [doi: 10.1126/science.286.5439.509] [Medline: 10521342]
34. Bjelland I, Krokstad S, Mykletun A, Dahl AA, Tell GS, Tambs K. Does a higher educational level protect against anxiety and depression? The HUNT study. Soc Sci Med 2008 Mar;66(6):1334-1345. [doi: 10.1016/j.socscimed.2007.12.019] [Medline: 18234406]
35. Elovainio M, Hakulinen C, Pulkki-Råback L, Juonala M, Raitakari OT. A network approach to the analysis of psychosocial risk factors and their association with health. J Health Psychol 2020 Sep;25(10-11):1587-1600. [doi: 10.1177/1359105318765624] [Medline: 29607696]

36. Ervasti J, Vahtera J, Pentti J, Oksanen T, Ahola K, Kivimäki M, et al. Depression-related work disability: socioeconomic inequalities in onset, duration and recurrence. PLoS One 2013;8(11):e79855 [FREE Full text] [doi: 10.1371/journal.pone.0079855] [Medline: 24278194]
37. Sobocki P, Ekman M, Agren H, Krakau I, Runeson B, Mårtensson B, et al. Health-related quality of life measured with EQ-5D in patients treated for depression in primary care. Value Health 2007;10(2):153-160 [FREE Full text] [doi: 10.1111/j.1524-4733.2006.00162.x] [Medline: 17391424]
38. Solmi M, Koyanagi A, Thompson T, Fornaro M, Correll CU, Veronese N. Network analysis of the relationship between depressive symptoms, demographics, nutrition, quality of life and medical condition factors in the Osteoarthritis Initiative database cohort of elderly North-American adults with or at risk for osteoarthritis – CORRIGENDUM. Epidemiol Psychiatr Sci 2019 Mar 08;29. [doi: 10.1017/s2045796019000064]
39. Jung JH, Seok H, Kim J, Song GG, Choi SJ. Association between osteoarthritis and mental health in a Korean population: a nationwide study. Int J Rheum Dis 2018 Mar;21(3):611-619. [doi: 10.1111/1756-185X.13098] [Medline: 28544532]
40. Semenkovich K, Brown ME, Svrakic DM, Lustman PJ. Depression in type 2 diabetes mellitus: prevalence, impact, and treatment. Drugs 2015 Apr;75(6):577-587. [doi: 10.1007/s40265-015-0347-4] [Medline: 25851098]
41. Wei Y, Cai D, Liu J, Liu R, Wang S, Tang Y, et al. Cholesterol and triglyceride levels in first-episode patients with major depressive disorder: A meta-analysis of case-control studies. J Affect Disord 2020 Apr 01;266:465-472. [doi: 10.1016/j.jad.2020.01.114] [Medline: 32056914]
42. Oh J, Kim T. Serum lipid levels in depression and suicidality: The Korea National Health and Nutrition Examination Survey (KNHANES) 2014. J Affect Disord 2017 Apr 15;213:51-58. [doi: 10.1016/j.jad.2017.02.002] [Medline: 28189965]
43. Sarris J, O'Neil A, Coulson CE, Schweitzer I, Berk M. Lifestyle medicine for depression. BMC Psychiatry 2014 Apr 10;14:107 [FREE Full text] [doi: 10.1186/1471-244X-14-107] [Medline: 24721040]
44. Simerville JA, Maxted WC, Pahira JJ. Urinalysis: a comprehensive review. Am Fam Physician 2005 Mar 15;71(6):1153-1162 [FREE Full text] [Medline: 15791892]
45. Ellis GG, Coppen A, Glen AI. Urine concentration in depressive illness. J Neurol Neurosurg Psychiatry 1971 Feb;34(1):30-31 [FREE Full text] [doi: 10.1136/jnnp.34.1.30] [Medline: 5551692]

Abbreviations

AUC: area under the receiver operating characteristic curve
EBICglasso: extended Bayesian information criterium graphical lasso
EQ-5D-3L: EuroQol five-dimension three-level
KCDC: Korea Centers for Disease Control and Prevention
KNHANES: Korea National Health and Nutrition Examination Survey
LASSO: least absolute shrinkage and selection operator
PHQ-9: Patient Health Questionnaire-9
PSU: primary sampling unit
QoL: quality of life
ROC: receiver operating characteristic
SES: socioeconomic status
TPOT: tree-based pipeline optimization tool
USG: urine specific gravity
XGBoost: extreme gradient boosting

Edited by R Kukafka; submitted 22.01.21; peer-reviewed by X Cheng, M Pradhan; comments to author 26.02.21; revised version received 06.03.21; accepted 06.05.21; published 24.06.21.

Please cite as:
Nam SM, Peterson TA, Seo KY, Han HW, Kang JI
Discovery of Depression-Associated Factors From a Nationwide Population-Based Survey: Epidemiological Study Using Machine Learning and Network Analysis
J Med Internet Res 2021;23(6):e27344
URL: https://www.jmir.org/2021/6/e27344/
doi:10.2196/27344
PMID:34184998

Original Paper

Chatbots to Support People With Dementia and Their Caregivers: Systematic Review of Functions and Quality

Nicole Ruggiano[1*], MSW, PhD; Ellen L Brown[2*], MS, RN, EdD, FAAN; Lisa Roberts[3*], PT, MS, DPT, CSCS; C Victoria Framil Suarez[4*], APRN, ANP-BC, DNP; Yan Luo[1*], BSW; Zhichao Hao[1*], MSW; Vagelis Hristidis[5*], PhD

[1]School of Social Work, University of Alabama, Tuscaloosa, AL, United States
[2]Graduate Nursing Department, Nicole Wertheim College of Nursing and Health Sciences, Florida International University, Miami, FL, United States
[3]Department of Physical Therapy, Nicole Wertheim College of Nursing and Health Sciences, Florida International University, Miami, FL, United States
[4]Department of Nursing, Nicole Wertheim College of Nursing and Health Sciences, Florida International University, Miami, FL, United States
[5]Department of Computer Science and Engineering, University of California, Riverside, Riverside, CA, United States
[*]all authors contributed equally

Corresponding Author:
Vagelis Hristidis, PhD
Department of Computer Science and Engineering
University of California, Riverside
317 Winston Chung Hall
900 University Ave
Riverside, CA, 92521
United States
Phone: 1 951 827 2478
Email: vagelis@cs.ucr.edu

Abstract

Background: Over the past decade, there has been an increase in the use of information technologies to educate and support people with dementia and their family caregivers. At the same time, chatbot technologies have become increasingly popular for use by the public and have been identified as having benefits for health care delivery. However, little is known about how chatbot technologies may benefit people with dementia and their caregivers.

Objective: This study aims to identify the types of current commercially available chatbots that are designed for use by people with dementia and their caregivers and to assess their quality in terms of features and content.

Methods: Chatbots were identified through a systematic search on Google Play Store, Apple App Store, Alexa Skills, and the internet. An evidence-based assessment tool was used to evaluate the features and content of the identified apps. The assessment was conducted through interrater agreement among 4 separate reviewers.

Results: Of the 505 initial chatbots identified, 6 were included in the review. The chatbots assessed varied significantly in terms of content and scope. Although the chatbots were generally found to be easy to use, some limitations were noted regarding their performance and programmed content for dialog.

Conclusions: Although chatbot technologies are well established and commonly used by the public, their development for people with dementia and their caregivers is in its infancy. Given the successful use of chatbots in other health care settings and for other applications, there are opportunities to integrate this technology into dementia care. However, more evidence-based chatbots that have undergone end user evaluation are needed to evaluate their potential to adequately educate and support these populations.

(J Med Internet Res 2021;23(6):e25006) doi:10.2196/25006

KEYWORDS

dementia; caregivers; chatbots; conversation agents; mobile apps; mobile phone

Introduction

Background

Over the past decade, chatbot technologies have increasingly been used by people to meet a variety of daily needs, including social and emotional support and information seeking [1]. The term *chatbot* refers to technologies that facilitate human-computer interaction by mimicking natural language conversations with users, either through text or spoken words [2]. Chatbots are also referred to as *conversation agents*, *dialog assistants*, *or intelligent virtual assistants*, the latter often referring to Google Assistant, iPhone *Siri*, and Amazon *Alexa*. Although interest in chatbots has increased among researchers, software developers, and end users alike, little is known about the efficacy of using chatbots to support people with dementia (which includes Alzheimer disease and related dementias) and or their caregivers. To address this gap in scholarship, the aim of this paper is to systematically review and evaluate the characteristics of chatbots that are currently available on the market and focus on dementia. The findings have implications for advancing technologies that support dementia care and caregiving, which has been identified as a top research priority [3].

Technologies for People With Dementia and Dementia Caregivers

There has been a growing interest in developing and testing technologies that can improve the quality of life and care for people with dementia and their family caregivers. Indeed, in 2018, the Family Caregiving Institute held a Research Priorities in Caregiving Summit: Advancing Family-Centered Care Across the Trajectory of Serious Illness. Among the 10 research priorities identified, the first 2 are “evaluate technologies that facilitate choice and shared decision making” and “determine where technology is best integrated across the trajectory of caregiving” [3].

Previous research has found that people with dementia have the ability to use a number of technologies designed to fit their needs, including computers [4] and touchscreen technologies [5,6]. Similarly, over the last decade, there has been increasing effort to develop technology-based interventions for dementia caregivers [7]. There is also a growing number of smartphone apps that target the needs of people with dementia and/or dementia caregivers [8,9]. In their systematic review of computer-based cognitive interventions for people with dementia, García-Casal et al [4] concluded that computer-based interventions demonstrated greater efficacy in improving cognition than noncomputer-based interventions. In a systematic review of technology-based interventions for dementia caregivers, it was reported that technology-based interventions often demonstrated efficacy in improving psychosocial outcomes but did not demonstrate efficacy in improving caregiving skills or care self-efficacy. Little is known about the use of chatbots among people with dementia and their caregivers [7].

Chatbot Technology and Health

Chatbots may use a variety of platforms, including websites, smartphones (eg, Siri), mobile apps (specialized in a domain such as depression or for general purpose, such as WhatsApp and Messenger), SMS text messaging, and in-home smart technologies (eg, Alexa) chatbots. Loranjo et al [10] identified 3 approaches for chatbots to manage dialog: (1) finite state, where there is a predetermined sequence of dialog steps that the user is led through; (2) frame based, where the chatbot asks the user questions and the user responses guide the chatbot through a flow of communication that is not predetermined; and (3) agent based, where artificial intelligence allows the chatbot and user to engage in complex communication. The same review reported that chatbot programs may allow dialog to be led by the chatbot system or the human user; dialog may either be in text or verbal formats.

One of the first established chatbots was programmed in the 1960s, called ELIZA, which was used for text-based psychiatric interviews [11]. This sparked the development of several early chatbots, such as GURU [12], CONVERSE [13], ALICE, and Elizabeth [14]. The renewed interest in chatbots in recent years is largely because of the advancements in deep learning technology [15], which enables agent-based chatbots, and the advancements in voice recognition technology, which have facilitated voice assistants such as Apple’s Siri, Microsoft’s Cortana, Amazon’s Alexa, and Google’s new Assistant [16].

Having been used for customer service, education, website user support, and entertainment, chatbots were found to have promising use in health care [10,17], especially for symptom self-assessment (eg, Babylon Health [18], Sensely, Ada Health, and your.md), and telemedicine [19,20]. Over time, chatbots have been developed and tested to address a number of health conditions and topics, including HIV and sexual health [21], substance abuse, assessment, mental health assessment [22], and weight management [23]. Overall, research on health-related chatbots has demonstrated promising results. In a review of chatbots for psychiatric care, it was reported that chatbots demonstrate efficacy in facilitating psychoeducation and self-adherence to treatment [24].

The benefits of using chatbots for health care have been identified for patients and health care systems alike. For example, chatbots may support patients in treatment management and moral support [25]. For health care systems using telehealth, well-designed chatbots can be a cost-effective way to collect patient data, facilitate patient education, improve engagement with patients, and allocate resources more efficiently [19]. A recent systematic review by Laranjo et al [10] concluded that despite the potential benefits of conversational agents in the health care field, limited research has evaluated the efficacy and safety of this technology.

Chatbots for Aging and Health

As the state-of-the-art advances, more older adults are using information technologies to meet a number of chronic health needs, such as increasing health self-efficacy, supporting self-care management, engaging in health promotion, and interacting with health care providers [26]. Previous studies have highlighted that auditory chatbots may be especially useful for older adults for health-related communication and information seeking because they operate through voice-driven conversation, which may be helpful for those with low computer literacy [27,28]. Chatbots, such as virtual assistants, have been used by older adults for many purposes. A previous study showed that voice-activated personal assistants were used as medication reminders for older adults [28]. Another study also indicated the application of chatbots to follow up older patients with cancer receiving chemotherapy at home [25].

In the case of dementia care and caregiving, information technologies have specifically been identified as a potential way to overcome the existing challenges of accessing education and emotional support [29,30]. Caregivers of people with dementia often experience depression, burden, and poor health outcomes because of lack of emotional and caregiving support, knowledge, and coping strategies for the complexity of dementia care [31]. This can have a negative impact on the quality of care and outcomes for people with dementia [31]. Information technology can overcome existing barriers to education and support for caregivers who are unable to attend in-person services or need flexible scheduling of services [8]. Preliminary studies on information and other mobile technologies have also demonstrated positive benefits for caregivers [8,32]. Mobile information technologies targeting people with dementia have mostly focused on improving memory and engagement [5]. Given the previous evidence presented on the benefits of using chatbots among older adults and in health care in general, they may pose unique benefits in providing education and support to people with dementia and their caregivers. They may especially be beneficial for those living in rural communities, who may face the challenge of distance and limited resources for interacting with providers and services [33]. It should be noted that there are commercially available apps that use chatbots for making clinical assessments of people with dementia [34]. However, many of these are designed to be used with the assistance of a professional and are not focused on education and support for care at home.

Need for Critical Review

Many of the identified uses for chatbots, such as self-assessments and health education, are potentially helpful in educating and supporting people with dementia and their caregivers. Advancing such technologies for the purpose of dementia care is critical, given that states are projected to increase the incidence of neurodegenerative diseases by up to 33% between 2019 and 2025 [31]. However, little is known about how chatbots may be useful for dementia care and caregiving. To address this gap, this study systematically identified and evaluated existing chatbot apps that are available on the market to educate and support people with dementia and their caregivers.

Methods

Overview

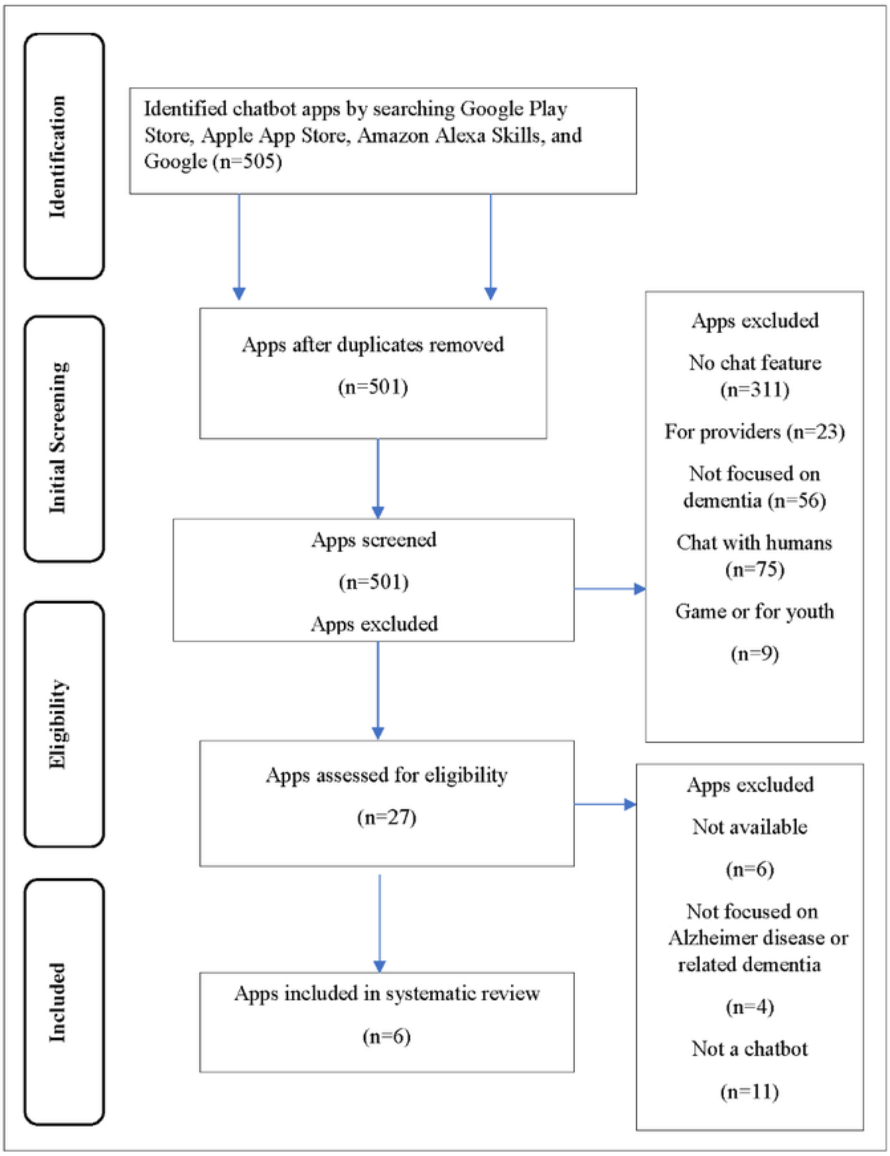
Previous studies on smartphone apps informed the methodology for evaluating existing chatbots [35]. A review of chatbot apps that focus on dementia was conducted between December 2019 and April 2020. Given the lack of chatbot app research in the academic literature, 3 investigators (NR, YL, and ZH) independently conducted searches on Google Play Store, Apple App Store, Amazon's Alexa Skills, and the internet. Key search terms used alone or in combination included "Chatbot," "Virtual Assistant," "Alzheimer's,"" dementia," "health," and "caregiver."

Chatbot apps were included for evaluation if they were programmed to engage in interactive dialog with the technology, verbally or through text; were focused on the topic of dementia; were designed for use by dementia caregivers and/or people with dementia; were available for download and use; were available in English; and were not games. Chatbots were excluded if they were programmed so that users chat with a live person, did not have content related to dementia, were designed to be used by providers, were not currently available for use, were not available in English, or were games. A total of 6 investigators were involved in evaluating apps for inclusion or exclusion criteria, and at least two investigators reviewed each app independently.

Critical Review

The evaluation tool of identified chatbots was developed using the framework developed by Radziwill and Benton [36] for evaluating chatbots, which is grounded in the computer science literature. The characteristics assessed are shown in Figure 1 and include *efficiency*, which included the category of *productivity*; *effectiveness*, which included the categories of *functionality* and *humanity*; and *satisfaction*, which included the categories of *affect* and *ethics and behavior.* A total of 3 authors (ELB, LR, and CVFS) evaluated each item of the assessment and graded the item using the following scale: ✓ (check), indicating that the evaluators could observe the attribute and/or the app-performed attribute function; — (em dash), indicating that the attribute or function was nonfunctional or inconsistent in functionality when accessed, and N/A (not applicable), indicating that the attribute did not apply to the stated functions of the app. The authors then compared their grades for each item and discussed discrepancies until an agreement was reached. A thematic analysis then was conducted based on the trend in scoring on the assessment across chatbots. All observations made by the team about the chatbot were based on their performance when evaluated in April 2020.

Figure 1. Flowchart of the chatbot apps searched and the screening procedures conducted using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.



Results

Overview

The results of the search strategy are shown in Figure 1. Overall, 6 chatbot apps targeted the topic of dementia: 1 text-based mobile app and 5 Alexa Skills voice apps. It should be noted that the text-based mobile app (CogniCare, CogniHealth Ltd) was associated with one of the Alexa Skills apps, and these apps were the only ones identified to have an associated website. Half of the apps focused on education about dementia, whereas the other half focused on memory and reminiscence. The educational content provided through the apps was generally about the epidemiology and symptoms of dementia and less focused on caregiving skills and activities. None of the apps made claims about the expected outcomes of using the app, and no peer-reviewed research was identified for any of the apps. General information about the apps reviewed can be found in Table 1, and the assessment of the apps can be found in Table 2. The assessment of the apps took place in May 2020; therefore, the characteristics of the app presented here may have changed since then.

Table 1. General information about the chatbot apps.

App name	App ID	Developer name	Operating system	Downloads (N)	Price	Privacy statement
CogniCare mobile app	1	CogniHealth Ltd (in partnership with Alzheimer Scotland)	Android and iOS	≥1000	Free	Yes
CogniCare (Alexa Skills version)	2	CogniHealth Ltd (in partnership with Alzheimer Scotland)	Alexa Skills	Not reported	Free	Yes
My Life Story	3	Z55-Studios	Alexa Skills	Not reported	Free	Yes
Dementia Types	4	T Tillman	Alexa Skills	Not reported	Free	Yes
Build Your Brain Power	5	Fire Up Your Goals LLC	Alexa Skills	Not reported	Free	Yes
Everything Memory	6	jaimiles23	Alexa Skills	Not reported	Free	Yes

Table 2. Assessment tool informed by the framework developed by Radziwill and Benton and evaluation of apps by the research team.

Category, Quality attribute[a]	App IDs					
	1	2	3	4	5	6
Efficiency						
Performance						
Robustness to unexpected input	✓[b]	—[c]	✓	—	—	—
Avoids inappropriate utterances and is able to perform damage control	N/A[d]	—	✓	—	—	N/A
Effective function allocation and provides appropriate escalation channels to humans	—	N/A	✓	—	—	N/A
Effectiveness						
Functionality						
Accurate speech synthesis	N/A	—	✓	—	✓	✓
Interprets commands accurately	✓	—	✓	—	✓	✓
General ease of use	✓	—	✓	✓	✓	✓
Executes requested tasks	✓	—	—	✓	✓	✓
Contains breadth of knowledge and is flexible in interpreting it	✓	—	N/A	✓	✓	✓
Humanity						
Passes the Turing test	N/A	—	✓	—	—	—
Convincing, satisfying, and natural interaction	N/A	—	✓	—	—	✓
Able to respond to specific questions	✓	—	N/A	—	N/A	N/A
Able to maintain themed discussion	N/A	—	N/A	✓	N/A	N/A
Satisfaction						
Affect						
Provides greetings and conveys personality	N/A	✓	✓	✓	✓	✓
Gives conversational cues	N/A	—	N/A	✓	✓	✓
Exudes warmth and authenticity	N/A	✓	✓	✓	—	✓
Makes tasks more fun and interesting	✓	—	✓	✓	—	✓
Entertains and/or enables participant to enjoy the interaction	✓	—	✓	✓	—	✓
Ethics and behavior						
Ethics and cultural knowledge of users	✓	✓	—	—	—	—
Protects and respects privacy	✓	✓	✓	✓	—	✓
Sensitivity to safety and social concerns	✓	✓	—	—	—	✓
Trustworthiness (of information source)	✓	✓	—	—	—	—
Awareness of trends and social context (in language and conversation)	—	—	—	—	—	—

[a] Apps were evaluated in May 2020.

[b] Evaluators could observe the attribute and/or the app-performed attribute function, as described earlier.

[c] The attribute or function was nonfunctional or inconsistent in functionality when accessed.

[d] N/A: not applicable.

Efficiency in Performance

For all apps, their functions and features were described in detail on Google Play Store or Alexa Skills pages. A total of 2 apps (app number 4 and 6) had more detailed instructions for getting started with using the apps than the others. One app (app number 2) offered an extensive guide to using the app with Alexa, although this guide was found on the developer's website and was not linked or listed on Amazon.com, where the app could be found and accessed. On Amazon.com, there was only a list of voice commands in speech bubbles that suggested questions to ask the chatbot. This posed challenges for the evaluators who initially accessed the apps for evaluation. The mobile text–based chatbot (app number 1) was the easiest to get started.

The evaluators also faced linguistic challenges when initially accessing the functions of the voice chatbot apps. For instance, some instances of Alexa's voice recognition were not flexible with the user's pronunciation of commands or required a specific command to get started in the chatbots, which was not immediately understood by the researchers. This may also be the reason that for voice chatbots, the evaluators noted that they would perform some of their described features, but not others. Only 1 chatbot (number 3) performed well on all 3 performance attributes. This was an audio chatbot that went through a series of questions about the user's life history, where the information was programmed into the app and the chatbot would tell the user their life story on command. For 3 of the chatbots, not all performance attributes were applicable.

Effectiveness

Functionality

In general, the chatbots were evaluated more favorably for their functionality, compared with the other categories assessed. In total, 5 of the 6 chatbots were found to be easy to use once the researchers were able to get the programs started. Of the 5 chatbots designed to educate about dementia, 3 were found to have a breadth of knowledge and flexibility in interpreting information.

Humanity

All apps demonstrated the ability to interact with users through voice or text responses to users' input. Although users found the apps' responses to their input to be understandable and logical, their responses were perceived to be more automated than natural conversation. Only 1 app (app number 3) passed the Turing test, which is the extent to which one can tell the difference between conversing with a human and the chatbot. The researchers also found that the programmed dialog in all the chatbots was limited in scope, which made extended language exchange between the user and chatbot to also be constrained. For example, 1 chatbot (app number 4) provided a definition for each type of dementia, but there was no additional content for a true back-and-forth exchange.

Satisfaction

Affect

In general, the evaluators assessed the chatbots as having features aimed at increasing the user's enjoyment while using them. For instance, 1 app (number 3) played calm music in the background while the users accessed its other features. All audio chatbots provided greetings and/or conveyed personality. Only 3 chatbots appeared to provide conversational cues to facilitate the user's ongoing engagement. For example, 1 chatbot would instruct the user to "Say, 'tell me more'" after it completed its dialog (app number 4). A total of 3 apps were evaluated as having the potential to make caregiving tasks more enjoyable or interesting.

Ethics and Behavior

Safety and privacy protection were evaluated for chatbot apps. In total, 5 apps had privacy statements describing the terms of use and security of users' input. In all the privacy statements, the developer confirmed that none of the users' information would be shared with third parties, unless the user gave permission to do so. In total, 2 apps (app number 2 and 6) included statements in their app descriptions that the chatbot was not a substitute for medical care or advice. In terms of trustworthiness of information, 4 apps were created by private developers. Two of the chatbot apps (app number 1 and 2) were developed by the same entity, which partnered with the nongovernmental organization Alzheimer Scotland to develop both the apps. It should be noted that since the time of review of the chatbots, the organization has also partnered with the University of Edinburgh for development. Hence, these apps were viewed as having the highest perception of trustworthiness.

Although the apps had content and functions that were deemed interesting and useful, some of the limitations in programmed content (as described earlier) presented challenges in this category as well. For instance, none of the apps were available in languages other than English, although 1 app (number 2) listed that their app was available in English from 5 different countries. The evaluators were not able to assess whether this resulted in language differences (eg, idioms and pronunciations) that were country specific. None of the apps were programmed to respond to patterns or the social context of user input.

Discussion

Principal Findings

At the time of this review, it was found that commercially available chatbot apps for people with dementia and their caregivers appeared to be in an early stage, with only 6 identified apps that met the search criteria across Google Play Store, the Apple Store, Amazon Alexa Skills, and the internet. All but one of these chatbot apps was the Alexa Skills app. Although chatbots, and specifically intelligent voice assistants, have been suggested as a potential source of assistance for older adults in general [37], their assessment indicates that people with dementia and their caregivers may experience challenges in using them. For instance, most apps had limitations in their programmed content, which created challenges when evaluators did not use specific commands. This may create barriers for populations with cognitive impairment or populations that use English as a second language. It may also create challenges for those who are not already familiar with chatbot technologies. Hence, people with dementia and caregivers could become frustrated when chatbots do not function as expected.

Limited program content also created challenges in having extended and/or varied conversations between users and chatbots. Owing to the complexity of dementia and variations in symptoms, this could limit the amount of education and support provided to a diverse population of people with dementia and their caregivers. This could impede the use of chatbots for information seeking and support. However, the hardware required for intelligent voice assistant platforms, such as Alexa, is recognized as having potential benefits for people with dementia and their caregivers. For instance, Alexa Skills relies on voice for navigation, which may be easier for older adults than typing, and does not require technical accessories, such as a mouse or keyboard. Hence, it may be worthwhile to further investigate how chatbots can be designed with user-friendly features for older adults, especially those who may have cognitive or memory impairments.

Several of the chatbots reviewed included educational information, although references were not provided for the informational sources; therefore, it was unclear whether health-related information provided by the chatbots is evidence based or whether information has been updated as needed. Designers should consider providing and using peer-reviewed references used in programming chatbots. In addition, although chatbot apps appeared to focus on education about dementia and engagement, the apps did not offer much information on caregiving skills and activities. However, the guidebook found for the CogniCare Alexa Skills chatbot reported that such content was being planned for the future. Such content would increase the usefulness of the chatbots for caregivers, who often feel isolated and have low caregiving self-efficacy.

Humanity Versus Usefulness

We found that most chatbots did not pass the Turing test, which mainly means that they were unable to keep the conversation going while pretending to be human. We note that acting human may not be enough if the chatbot does not achieve its goals of educating or helping caregivers or patients. Most early research on chatbots has focused on the humanity aspect [11], whereas more recent research has focused on how to create goal-oriented chatbots, which understand the user and complete a transaction, such as making an appointment or answering a question. Much of this recent work assumes that there is a data set of conversation on the domain to train, which is not the case for dementia care, where there are no available conversations between patients and providers. Nevertheless, as more artificial intelligence tools are created, we expect that the responsiveness of dementia chatbots will improve as well.

Need for Evidence-Based Chatbots for Dementia Care and Caregiving

It is important to explore how chatbots may be used to educate and support people with dementia and their families. There are many potential chatbot functions that those affected by dementia may value depending on their preferences and needs. For instance, given the previous applications of chatbots to adult learning, chatbots could be used to educate and train caregivers on a variety of topics, especially those who feel isolated [38]. Chatbots can also be programmed so that caregivers can take self-assessments and receive evidence-based feedback or be linked with resources. For people with dementia, designers may consider how evidence-based cognitive training and other evidence-based programs can be administered via a chatbot [39].

When considering people with dementia as targeted end users, designers should also consider that there is a continuum of severity in dementia symptomology and assess how content and functions may be appropriate for different stages of disease progression or adaptable to be used as symptoms progress. For example, people with mild dementia may look for and benefit from features that support independence or memory-building activities. However, people in later stages of dementia and their caregivers may prefer features that promote engagement and reminiscence, such as a recorded message from a loved one or a favorite song. Listening to music can be a shared enjoyable activity by the care recipient and caregiver, and music therapy has been found to be therapeutic in managing behavioral disturbance in people with dementia [40].

Although there are opportunities to advance the state of the science on dementia-focused chatbots in general, it is important that such chatbots should be evidence based. It is unclear whether the chatbots included in this review underwent empirical user testing. It should be noted that since conducting the chatbot review, the developers of CogniCare have published a peer-reviewed article on the development of their apps [41], although this article did not report outcome findings for users. Overall, this underscores the importance of developing chatbot technologies that undergo extensive user testing, given that the technical literacy required to use chatbots can pose a barrier to older adults' use and acceptability [25]. This is very important for chatbots designed to be used within the context of health care, given the previously stated concerns that input inflexibility of a particular chatbot could pose risks to patients [10]. This point also relates to the quality of the content programmed in a chatbot. For instance, it is unclear the extent to which educational material provided through the chatbots assessed for this study were based on evidence from the literature on dementia.

Opportunities for Integrating Chatbots Into Dementia Care Delivery

There are particular potential benefits of integrating chatbot technology into dementia care settings. People with dementia and their caregivers are at an increased risk of social isolation, especially for people with dementia who are in long-term care settings. The potential for chatbots for socially isolated populations has been particularly evident during the COVID-19 pandemic, where the Center for Disease Control and Prevention launched its Clara Bot app as an interactive way to educate the public and provide emotional support [42]. Integrating chatbots into dementia clinical settings could also potentially reduce the burden on staff by offering vetted information to patient and caregiver populations, which could reduce calls and questions for clinical staff. It should be noted that the quality of the information provided by chatbots in this review was not fully evaluated, and it is unclear if they provided vetted information. In addition, the limitations of the chatbots critically reviewed in this study, such as the need for specific phrases and pronunciations, highlight the need for developing health care chatbots that support diverse patient populations. For instance, dementia chatbots that are flexible in their understanding of varying accents and dialects will increase their acceptability and ease of use.

A key obstacle in integrating chatbots into care delivery is the Health Insurance Portability and Accountability Act (HIPAA) compliance needed for the chatbot to be able to exchange information with a health care entity (covered entity). Fortunately, Amazon Alexa has recently begun offering HIPAA-compliant skills [43]. Mobile apps can also be HIPAA compliant, but this is not the case for chatbots deployed on existing platforms, such as Messenger, SMS text messaging, or WhatsApp [44].

Limitations and Future Needs

There are limitations to this critical review that should be considered when interpreting the results. Despite the extensive search and review process used to identify chatbots, it is possible that there may be additional chatbots that address dementia but were not identified. In addition, there are no standardized measurement tools for chatbot technology. The authors initially attempted to conduct this review using the widely used mobile app rating scale [45]. However, too many items did not apply to chatbots. It should be noted that the evaluation tool used for the review developed by Radziwill and Benton [36] is heavily grounded in evidence on chatbots.

Conclusions

Although chatbots offer the potential for an engaging technology that can benefit people with dementia and their caregivers, more work is needed to develop evidence-based chatbots that are easy to use for these populations. Future research in this area should involve interdisciplinary scientific teams that have the expertise to develop a high-performing chatbot technology as well as expertise in dementia or health education.

Conflicts of Interest

VH is the founder of SmartBot360, which is a health care chatbot company. SmartBot360 was not included in this survey, as it does not focus on dementia caregivers.

References

1. Brandtzaeg PB, Følstad A. Why people use chatbots. In: Internet Science. Switzerland: Springer; 2017:377-392.
2. Abdul-Kader SA, Woods J. Survey on chatbot design techniques in speech conversation systems. Int J Adv Comput Sci Appl 2015;6(7):A. [doi: 10.14569/IJACSA.2015.060712]
3. Harvath TA, Mongoven JM, Bidwell JT, Cothran FA, Sexson KE, Mason DJ, et al. Research priorities in family caregiving: process and outcomes of a conference on family-centered care across the trajectory of serious illness. Gerontologist 2020 Feb 14;60(Suppl 1):5-13 [FREE Full text] [doi: 10.1093/geront/gnz138] [Medline: 32057081]
4. García-Casal JA, Loizeau A, Csipke E, Franco-Martín M, Perea-Bartolomé MV, Orrell M. Computer-based cognitive interventions for people living with dementia: a systematic literature review and meta-analysis. Aging Ment Health 2017 Dec;21(5):454-467. [doi: 10.1080/13607863.2015.1132677] [Medline: 26806365]
5. Hitch D, Swan J, Pattison R, Stefaniak R. Use of touchscreen tablet technology by people with dementia in homes: a scoping review. J Rehabil Assist Technol Eng 2017;4:A [FREE Full text] [doi: 10.1177/2055668317733382] [Medline: 31186940]
6. Joddrell P, Astell AJ. Studies involving people with dementia and touchscreen technology: a literature review. JMIR Rehabil Assist Technol 2016 Nov 04;3(2):e10 [FREE Full text] [doi: 10.2196/rehab.5788] [Medline: 28582254]
7. Ruggiano N, Brown EL, Li J, Scaccianoce M. Rural dementia caregivers and technology: what is the evidence? Res Gerontol Nurs 2018 Jul 01;11(4):216-224. [doi: 10.3928/19404921-20180628-04] [Medline: 30036405]
8. Brown EL, Ruggiano N, Page TF, Roberts L, Hristidis V, Whiteman KL, et al. CareHeroes web and Android™ apps for dementia caregivers: a feasibility study. Res Gerontol Nurs 2016;9(4):193-203 [FREE Full text] [doi: 10.3928/19404921-20160229-02] [Medline: 29977440]
9. Guo Y, Yang F, Hu F, Li W, Ruggiano N, Lee HY. Correction: existing mobile phone apps for self-care management of people with Alzheimer disease and related dementias: systematic analysis. JMIR Aging 2020 May 19;3(1):e18754 [FREE Full text] [doi: 10.2196/18754] [Medline: 32427566]
10. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
11. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966 Jan;9(1):36-45. [doi: 10.1145/365153.365168]
12. Colby KM, Colby PM, Stoller RJ. Dialogues in natural language with GURU, a psychologic inference engine. Philos Psychol 2008 Jun 10;3(2-3):171-186. [doi: 10.1080/09515089008572997]
13. Batacharia B, Levy D, Catizone R, Krotov A, Wilks Y. Converse: a conversational companion. In: Machine Conversation. Boston: Springer; 1999:205-215.
14. Shawar BA, Atwell E. A comparison between Alice and Elizabeth chatbot systems. University of Leeds, School of Computing Research Report Series. 2002. URL: http://eprints.whiterose.ac.uk/81930/1/AComparisonBetweenAliceElizabeth.pdf [accessed 2011-05-11]
15. Britz D, Goldie A, Luong M, Le Q. Massive exploration of neural machine translation architectures. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017 Presented at: Conference on Empirical Methods in Natural Language Processing; September, 2017; Copenhagen, Denmark p. 1442-1451. [doi: 10.18653/v1/D17-1151]
16. Dale R. The return of the chatbots. Nat Lang Eng 2016 Sep 13;22(5):811-817 [FREE Full text] [doi: 10.1017/S1351324916000243]

17. Shawar BA, Atwell E. Different measurements metrics to evaluate a chatbot system. In: Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies. 2007 Presented at: Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies; April, 2007; Rochester, NY p. 89-96 URL: https://www.aclweb.org/anthology/W07-0313
18. Babylon. URL: https://www.babylonhealth.com/ [accessed 2021-05-19]
19. Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth '17: 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May, 2017; Barcelona Spain p. 261-265. [doi: 10.1145/3154862.3154914]
20. Fadhil A. CoachAI: A Conversational-UI Assisted E-coaching platform For Health & Wellbeing 2019:A. [doi: 10.13140/RG.2.2.22231.32164]
21. Brixey J, Hoegen R, Lan W, Rusow J, Singla K, Yin X, et al. SHIHbot: a Facebook chatbot for sexual health information on HIV/AIDS. In: Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue. 2017 Presented at: 18th Annual SIGdial Meeting on Discourse and Dialogue; August, 2017; Saarbrücken, Germany p. 370-373. [doi: 10.18653/v1/w17-5544]
22. Elmasri D, Maeder A. A conversational agent for an online mental health intervention. In: Brain Informatics and Health. Switzerland: Springer; 2016:243-251.
23. Huang CY, Yang MC, Huang CY, Chen YJ, Wu ML, Chen KW. A chatbot-supported smart wireless interactive healthcare system for weight control and health promotion. In: Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). 2018 Presented at: IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); Dec. 16-19, 2018; Bangkok, Thailand. [doi: 10.1109/ieem.2018.8607399]
24. 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]
25. Piau A, Crissey R, Brechemier D, Balardy L, Nourhashemi F. A smartphone Chatbot application to optimize monitoring of older patients with cancer. Int J Med Inform 2019 Aug;128:18-23. [doi: 10.1016/j.ijmedinf.2019.05.013] [Medline: 31160007]
26. Portz J, Miller A, Foster B, Laudeman L. Persuasive features in health information technology interventions for older adults with chronic diseases: a systematic review. Health Technol 2016 Mar 23;6(2):89-99 [FREE Full text] [doi: 10.1007/s12553-016-0130-x]
27. Brewer RN, Findlater L, Kaye JJ, Lasecki W, Munteanu C, Weber A. Accessible voice interfaces. In: Proceedings of the Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2018 Presented at: CSCW '18: Computer Supported Cooperative Work and Social Computing; November, 2018; Jersey City NJ USA p. 441-446. [doi: 10.1145/3272973.3273006]
28. Dahlke DV, Ory MG. Emerging opportunities and challenges in optimal aging with virtual personal assistants. Pub Pol Aging Rep 2017;27(2):68-73. [doi: 10.1093/ppar/prx004]
29. Vedel I, Akhlaghpour S, Vaghefi I, Bergman H, Lapointe L. Health information technologies in geriatrics and gerontology: a mixed systematic review. J Am Med Inform Assoc 2013 Nov;20(6):1109-1119 [FREE Full text] [doi: 10.1136/amiajnl-2013-001705] [Medline: 23666776]
30. Zulman DM, Piette JD, Jenchura EC, Asch SM, Rosland A. Facilitating out-of-home caregiving through health information technology: survey of informal caregivers' current practices, interests, and perceived barriers. J Med Internet Res 2013 Jul 10;15(7):e123 [FREE Full text] [doi: 10.2196/jmir.2472] [Medline: 23841987]
31. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. Alzheimers Dement 2019 Mar 01;15(3):321-387. [doi: 10.1016/j.jalz.2019.01.010]
32. Sikder AT, Yang FC, Schafer R, Dowling GA, Traeger L, Jain FA. Mentalizing imagery therapy mobile app to enhance the mood of family dementia caregivers: feasibility and limited efficacy testing. JMIR Aging 2019 Mar 21;2(1):e12850 [FREE Full text] [doi: 10.2196/12850] [Medline: 31518275]
33. O'Connell ME, Crossley M, Cammer A, Morgan D, Allingham W, Cheavins B, et al. Development and evaluation of a telehealth videoconferenced support group for rural spouses of individuals diagnosed with atypical early-onset dementias. Dementia (London) 2014 May;13(3):382-395. [doi: 10.1177/1471301212474143] [Medline: 24339063]
34. Thabtah F, Peebles D, Retzler J, Hathurusingha C. Dementia medical screening using mobile applications: a systematic review with a new mapping model. J Biomed Inform 2020 Nov;111:103573 [FREE Full text] [doi: 10.1016/j.jbi.2020.103573] [Medline: 32961306]
35. Brown EL, Ruggiano N, Li J, Clarke PJ, Kay ES, Hristidis V. Smartphone-based health technologies for dementia care: opportunities, challenges, and current practices. J Appl Gerontol 2019 Jan;38(1):73-91. [doi: 10.1177/0733464817723088] [Medline: 28774215]
36. Radziwill NM, Benton MC. Evaluating quality of chatbots and intelligent conversational agents. arXiv - Computer Science, Computers and Society. 2017. URL: https://arxiv.org/abs/1704.04579 [accessed 2011-05-11]

37. Reis A, Paulino D, Paredes H, Barroso I, Monteiro M, Rodrigues V, et al. Using intelligent personal assistants to assist the elderlies: an evaluation of Amazon Alexa, Google Assistant, Microsoft Cortana, and Apple Siri. In: Proceedings of the 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW). 2018 Presented at: 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW); June 20-22, 2018; Thessaloniki, Greece p. 1-5. [doi: 10.1109/tishw.2018.8559503]
38. Huang W. Designing and evaluating three chatbot-enhanced activities for a flipped graduate course. Int J Mech Eng Robot Res 2019:813-818. [doi: 10.18178/ijmerr.8.5.813-818]
39. Smith GE, Housen P, Yaffe K, Ruff R, Kennison RF, Mahncke HW, et al. A cognitive training program based on principles of brain plasticity: results from the Improvement in Memory with Plasticity-based Adaptive Cognitive Training (IMPACT) study. J Am Geriatr Soc 2009 Apr;57(4):594-603 [FREE Full text] [doi: 10.1111/j.1532-5415.2008.02167.x] [Medline: 19220558]
40. Ueda T, Suzukamo Y, Sato M, Izumi S. Effects of music therapy on behavioral and psychological symptoms of dementia: a systematic review and meta-analysis. Ageing Res Rev 2013 Mar;12(2):628-641 [FREE Full text] [doi: 10.1016/j.arr.2013.02.003] [Medline: 23511664]
41. Macaden L, Muirhead K, Melchiorre G, Mantle R, Ditta G, Giangreco A. Relationship-centred CogniCare: an academic–digital–dementia care experts interface. Working with Older People 2020 Dec 24;25(1):73-83 [FREE Full text] [doi: 10.1108/WWOP-05-2020-0016]
42. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. NPJ Digit Med 2020;3:65 [FREE Full text] [doi: 10.1038/s41746-020-0280-0] [Medline: 32377576]
43. Jiang R. Introducing new Alexa healthcare skills. Amazon Alexa. 2019 Apr 4. URL: https://developer.amazon.com/blogs/alexa/post/ff33dbc7-6cf5-4db8-b203-99144a251a21/introducing-new-alexa-healthcare-skills [accessed 2011-05-11]
44. Hristidis V. When is a chatbot HIPAA-compliant? Medium. 2018. URL: https://medium.com/@vagelis/when-is-a-chatbot-hipaa-compliant-bb1862270da4 [accessed 2011-05-11]
45. 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;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act

Edited by R Kukafka; submitted 14.10.20; peer-reviewed by F Jain, P Zou, Y Guo; comments to author 23.12.20; revised version received 23.02.21; accepted 25.04.21; published 03.06.21.

Please cite as:
Ruggiano N, Brown EL, Roberts L, Framil Suarez CV, Luo Y, Hao Z, Hristidis V
Chatbots to Support People With Dementia and Their Caregivers: Systematic Review of Functions and Quality
J Med Internet Res 2021;23(6):e25006
URL: https://www.jmir.org/2021/6/e25006
doi:10.2196/25006
PMID:34081019

Original Paper

Development of a Positive Body Image Chatbot (KIT) With Young People and Parents/Carers: Qualitative Focus Group Study

Francesca Beilharz[1], DPsych; Suku Sukunesan[2], PhD; Susan L Rossell[3,4], PhD; Jayashri Kulkarni[1], MBBS, PhD; Gemma Sharp[1], PhD

[1]Monash Alfred Psychiatry Research Centre, Monash University, Melbourne, Australia
[2]Swinburne Business School, Swinburne University of Technology, Melbourne, Australia
[3]Centre for Mental Health, Swinburne University of Technology, Melbourne, Australia
[4]Psychiatry, St Vincent's Hospital, Melbourne, Australia

Corresponding Author:
Gemma Sharp, PhD
Monash Alfred Psychiatry Research Centre
Monash University
4/607 St Kilda Road
Melbourne, 3004
Australia
Phone: 61 390765167
Email: gemma.sharp@monash.edu

Abstract

Background: Body image and eating disorders represent a significant public health concern; however, many affected individuals never access appropriate treatment. Conversational agents or chatbots reflect a unique opportunity to target those affected online by providing psychoeducation and coping skills, thus filling the gap in service provision.

Objective: A world-first body image chatbot called "KIT" was designed. The aim of this study was to assess preliminary acceptability and feasibility via the collection of qualitative feedback from young people and parents/carers regarding the content, structure, and design of the chatbot, in accordance with an agile methodology strategy. The chatbot was developed in collaboration with Australia's national eating disorder support organization, the Butterfly Foundation.

Methods: A conversation decision tree was designed that offered psychoeducational information on body image and eating disorders, as well as evidence-based coping strategies. A version of KIT was built as a research prototype to deliver these conversations. Six focus groups were conducted using online semistructured interviews to seek feedback on the KIT prototype. This included four groups of people seeking help for themselves (n=17; age 13-18 years) and two groups of parents/carers (n=8; age 46-57 years). Participants provided feedback on the cartoon chatbot character design, as well as the content, structure, and design of the chatbot webchat.

Results: Thematic analyses identified the following three main themes from the six focus groups: (1) chatbot character and design, (2) content presentation, and (3) flow. Overall, the participants provided positive feedback regarding KIT, with both young people and parents/carers generally providing similar reflections. The participants approved of KIT's character and engagement. Specific suggestions were made regarding the brevity and tone to increase KIT's interactivity.

Conclusions: Focus groups provided overall positive qualitative feedback regarding the content, structure, and design of the body image chatbot. Incorporating the feedback of lived experience from both individuals and parents/carers allowed the refinement of KIT in the development phase as per an iterative agile methodology. Further research is required to evaluate KIT's efficacy.

(J Med Internet Res 2021;23(6):e27807) doi:10.2196/27807

KEYWORDS

body image; eating disorder; chatbot; conversational agent; artificial intelligence; mental health; digital health; design

Introduction

Background

Mental disorders are a major public health concern, with a 29% lifetime prevalence rate across the general population [1]. In addition to reduced quality of life and individual impact, mental disorders have a significant global economic cost, which has been predicted at US $16.3 trillion worldwide from 2011 to 2030 [2]. Eating disorders have the highest mortality rate of any mental health diagnosis [3,4], and global prevalence rates have doubled from 3.4% to 7.8% of the population from 2000 to 2018 [5]. Despite the significant physical, mental, and social consequences associated with eating disorders, more than 75% of these individuals do not access appropriate treatment [6].

A major risk factor for developing eating disorders or other serious mental health conditions is body image dissatisfaction or concern [7]. This risk is increased by using social media, particularly image-based platforms such as Facebook, Instagram, and TikTok [8]. For example, using social media for 30 minutes a day, particularly looking at photos of peers and celebrities, has been shown to negatively impact women's body image and mood [9]. Over 3.6 billion people worldwide used social media in 2020, with many young people spending a significantly increasing amount of time on these platforms [10]. Consequently, social media, where young people are spending their time, presents a unique pathway to prevent and intervene with body image and eating concerns.

Guidelines for the evidence-based treatment of eating disorders recommend prevention and early intervention strategies for the best prognosis, with most individuals developing these concerns as young people [7,11]. Therefore, there is a need for early intervention and support delivery that is accessible and available to young people. The therapeutic approach with the strongest evidence base across the full spectrum of eating disorder presentations is enhanced cognitive behavioral therapy (CBT-E) where the core psychopathology addressed in treatment is the overevaluation of weight and shape (broadly, negative body image) [12]. Psychoeducation, or communicating relevant information about mental health conditions and treatments, is deemed an essential part of all psychological treatments, including CBT-E [12,13], especially given the wealth of misinformation in the general community about "healthy" eating and weight [11]. Psychoeducation is associated with small yet significant improvements in body image, which is further enhanced when combined with additional interventions [14,15].

The development of alternative coping skills is another requirement of evidence-based treatments [16], such as mindfulness practice, which involves the nonjudgemental awareness and acceptance of the present moment. Mindfulness training has been shown to significantly improve body image dissatisfaction and can effectively be delivered online as brief microinterventions [17-19]. When combined with psychoeducation, these brief evidence-based interventions (microinterventions) are likely to be beneficial in reducing body image concerns. Therefore, providing an accessible platform where individuals and support people can access evidence-based psychoeducation and coping skills may assist in addressing negative body image, and, in turn, potentially intervene in the development of eating disorders.

Prior Work

Digital mental health interventions have been increasing in popularity and efficacy, with chatbots being one such format. A chatbot allows the provision of information and services through a message-based interface, which is available on messaging platforms, websites, and apps. The conversational language and format of chatbots are relevant in connecting with youth, and are a promising avenue to supplement traditional support services [20]. Chatbots in the mental health sphere offer several benefits, including wider accessibility, instantaneous responses, and low or no cost to users, and can act as a stepping stone to link with more individualized or alternative services [21]. An additional barrier to accessing traditional services can be mental health stigma or lack of available services, where individuals may feel uncomfortable or unable to reach out for help. Digital mental health services can help bridge this gap, as anonymously conversing with artificial intelligence can feel less intimidating, particularly for young people who have grown up in the digital age [21].

The evidence for chatbots as supplements to mental health support is growing, with generally positive outcomes [20,22]. A recent systematic review and meta-analysis of 12 studies suggested that while heterogeneity among chatbot studies limits the interpretations of pooled efficacy, individual studies have evidenced improvements in symptoms of depression, anxiety, and general coping skills [23]. Owing to the limited conclusions that can be made, the authors recommend that chatbots be used to supplement treatment from mental health professionals rather than as a replacement. A scoping review of 41 chatbots reported that the most common uses of chatbots are therapy provision, information delivery, and screening, with the majority focused on mood and anxiety disorders [24]. However, to the authors' knowledge, there are no chatbots currently available that specifically target the increasingly common issue of negative body image. Given the rise of social media leading to increased appearance comparisons, "proeating disorder" content online, and "health" misinformation, there is a need for a reliable, evidence-based, and supportive digital mental health service within the field so users can receive more timely support for body image concerns.

Our Study

Based on this gap in service provision, we initiated the development of a novel body image chatbot. The chatbot provides psychoeducation and coping skills targeting body image concerns, for the purposes of prevention and supplementing traditional forms of treatment. This study aimed to assess the preliminary acceptability and feasibility of such a chatbot by collecting qualitative data from focus groups. This was an early stage evaluation, predominantly focused on the content, structure, and design of the chatbot in order to refine the design prior to launch, and the assessment of efficacy.

Methods

KIT Intervention

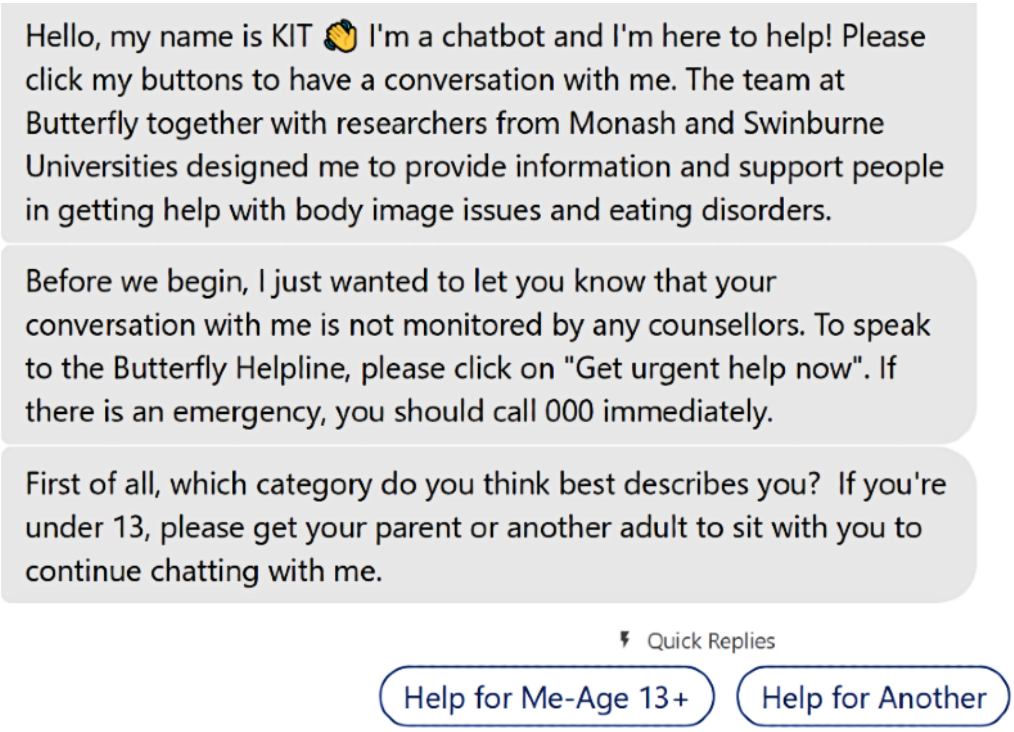
The chatbot, called “KIT,” is a conversational agent designed to support people with concerns surrounding body image and eating issues, as well as their loved ones (eg, parents, partners, and friends). The preliminary conversational content and decision tree for KIT was developed by the authors, in collaboration with the helpline and communications teams at the Butterfly Foundation (a national charity based in Australia, which provides support for body image and eating disorders). The conversation content was based on evidence-based information/interventions for eating disorders, specifically, psychoeducation, cognitive behavioral therapy (CBT), acceptance commitment therapy (ACT), and mindfulness [25,26], and adapted for delivery by a chatbot. Owing to the short and simple style of conversations KIT was designed to deliver, we were highly selective in the therapeutic elements we chose from CBT (eg, education on cognitive distortions or unhelpful thinking styles), ACT (eg, practicing detaching from unhelpful thoughts via cognitive defusion exercises), and mindfulness (eg, mindful breathing) [17,25,26].

In an effort to enhance the co-design of KIT, preliminary online forums were held with young people aged 13 to 18 years who self-identified as having lived experience with body image concerns and/or eating disorders. Please note that the reason this age range was chosen was to ensure that KIT’s dialog was understandable to the youngest users possible, not that KIT’s usage is limited to people aged 13 to 18 years. We also ran preliminary online forums with parents/carers of people with lived experience of body image concerns and/or eating disorders. The preliminary forums were crucial to assist with the refinement of the dialog, particularly devising shorter and more conversational ways of delivering educational information, as well as the design of the decision tree. The importance of involving end users in a software project has garnered strong support over recent years.

A professional chatbot production company, Proxima, using Iris Conversational Intelligence software [27], built the prototype chatbot based on the co-designed decision tree and conversation content. The prototype chatbot took the form of a webchat and was accessible on both mobile and desktop devices.

In terms of the specific chatbot content, each interaction with KIT began with a welcome message explaining the purpose and capabilities of the chatbot (Figure 1). This included directions to proceed through KIT by selecting buttons and information that the service was not monitored by any counsellors. Users were directed to speak to the Butterfly helpline or call emergency services if they required a higher level of assistance. KIT then asked the primary reason the user was there (“help for me [aged 13+]” or “help for another”). Once users were directed down one of these two pathways, they could self-direct to various psychoeducation or microintervention contents. The minimum age to use KIT was determined to be 13 years, as this is generally the age at which young people are permitted to sign up for a social media account [28].

Figure 1. KIT’s welcome message for users explaining capabilities and prompts for users to choose the “help for me” or “help for another” pathway.



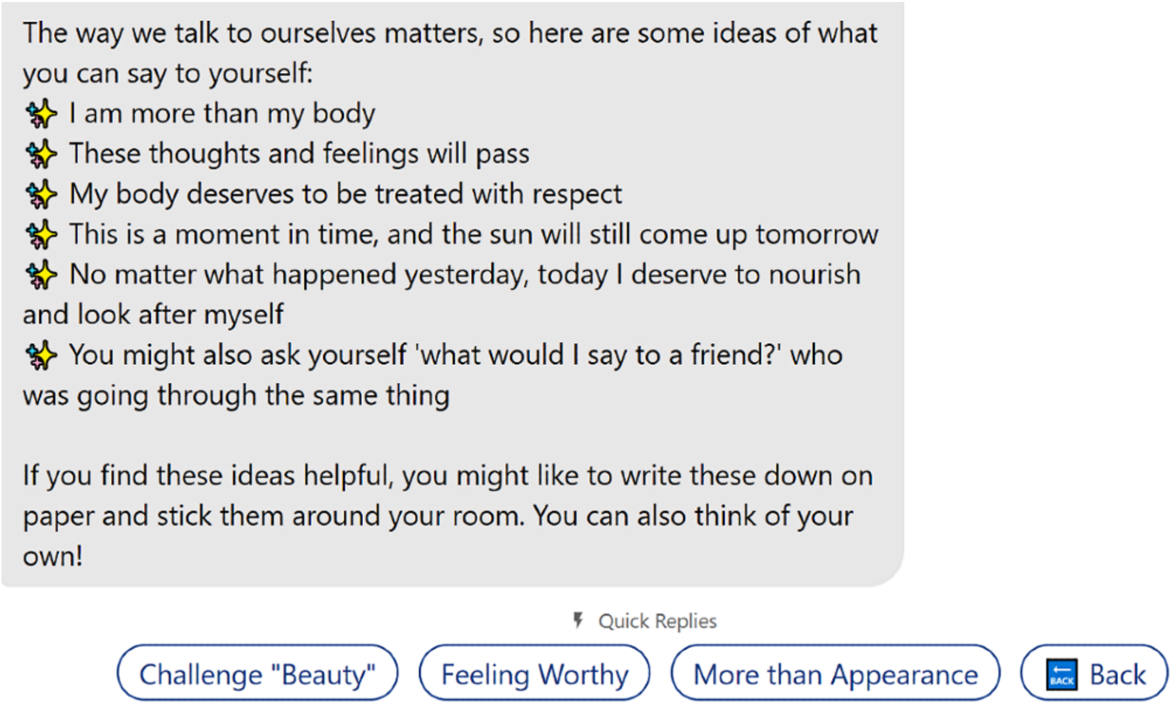
Hello, my name is KIT 👋 I'm a chatbot and I'm here to help! Please click my buttons to have a conversation with me. The team at Butterfly together with researchers from Monash and Swinburne Universities designed me to provide information and support people in getting help with body image issues and eating disorders.

Before we begin, I just wanted to let you know that your conversation with me is not monitored by any counsellors. To speak to the Butterfly Helpline, please click on "Get urgent help now". If there is an emergency, you should call 000 immediately.

First of all, which category do you think best describes you? If you're under 13, please get your parent or another adult to sit with you to continue chatting with me.

For people aged 13 years or above, KIT included psychoeducation conversations regarding body image, eating disorders and body dysmorphic disorders, health impacts, risk and protective factors, how to seek help, and support during the COVID-19 pandemic. The coping skills section covered strategies for managing social media, “busting myths” about food, comparisons and self-perception, challenging beauty ideals, mindfulness, managing difficult thoughts and emotions, body positive self-talk, enjoyable movement, and improving self-worth (Figure 2).

Figure 2. An example of one of KIT's coping skills conversations, "Positive Self-Talk," with prompts to continue the conversation with other skills that also aim to promote positive body image.

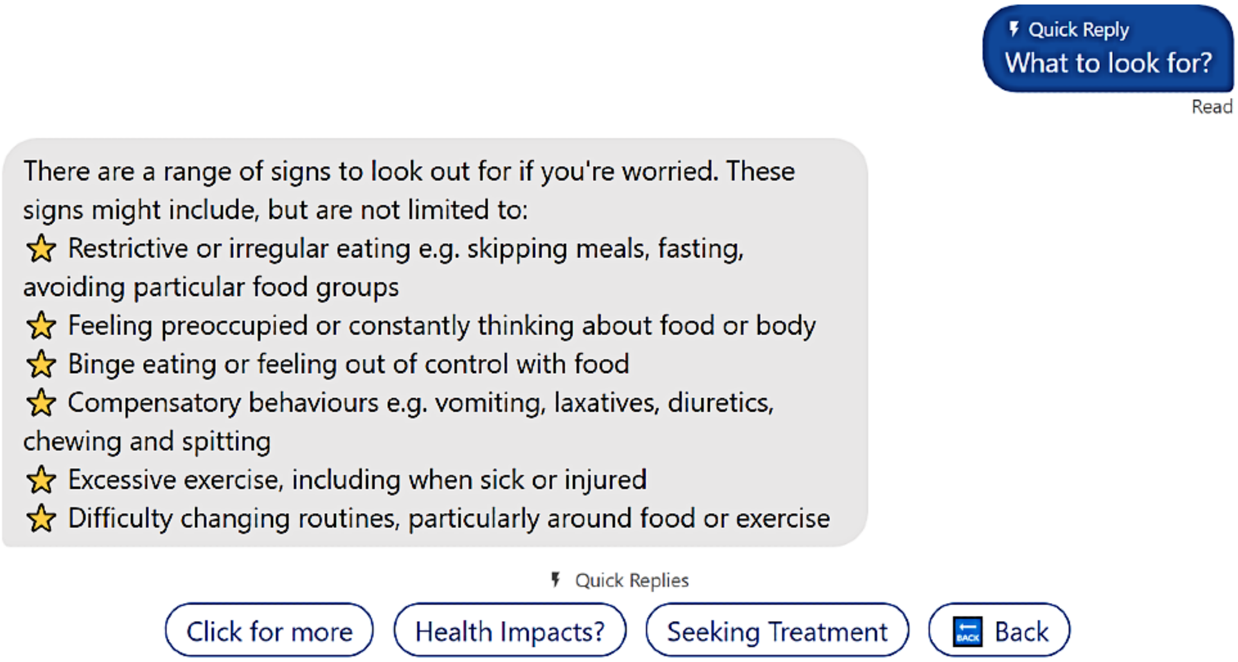


The way we talk to ourselves matters, so here are some ideas of what you can say to yourself:
✨ I am more than my body
✨ These thoughts and feelings will pass
✨ My body deserves to be treated with respect
✨ This is a moment in time, and the sun will still come up tomorrow
✨ No matter what happened yesterday, today I deserve to nourish and look after myself
✨ You might also ask yourself 'what would I say to a friend?' who was going through the same thing

If you find these ideas helpful, you might like to write these down on paper and stick them around your room. You can also think of your own!

The pathway for people seeking to help someone else explored similar psychoeducation conversations, with the addition of further information regarding warning signs and treatment pathways (Figure 3). An additional section was included where users could seek support for themselves, including individual support and information about support groups. For all users, KIT included a "get urgent help now" section that was linked to the Butterfly helpline and other crisis support services.

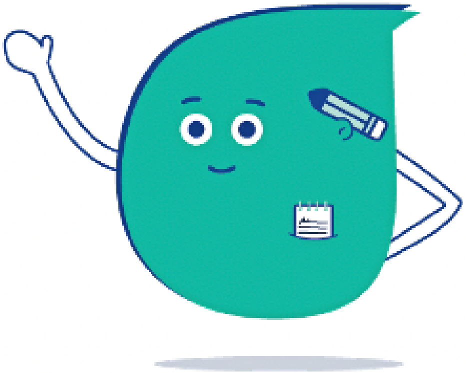
Figure 3. An example of one of KIT's psychoeducation conversations in the "help for another" pathway where the user has asked "What to look out for?" as in eating disorder warning signs. The "Click for more" option provides further warning signs.



There are a range of signs to look out for if you're worried. These signs might include, but are not limited to:
⭐ Restrictive or irregular eating e.g. skipping meals, fasting, avoiding particular food groups
⭐ Feeling preoccupied or constantly thinking about food or body
⭐ Binge eating or feeling out of control with food
⭐ Compensatory behaviours e.g. vomiting, laxatives, diuretics, chewing and spitting
⭐ Excessive exercise, including when sick or injured
⭐ Difficulty changing routines, particularly around food or exercise

To increase the appeal of the chatbot KIT service, a chatbot "character" (Figure 4) was designed by a professional graphic design company, Yoke, in collaboration with the authors and the helpline and communications teams at the Butterfly Foundation. Preliminary online forums were held with young people and parents/carers to inform the feature development of the character prior to the design by the graphic design company.

Figure 4. The main pose for the chatbot character of KIT.



Participants

Online advertisements were placed on Butterfly Foundation social media pages for people living in Australia (aged 13 to 18 years) with lived experience of body image concerns and/or eating disorders, and parents/carers of people with these lived experiences. Potential participants completed an online expression of interest form linked to the advertisement and were then invited (by the author GS or FB) to participate in the online focus group based on availability.

There were two focus groups for parents/carers, which included participants aged 46 to 57 years, with two males and six females, evenly split into groups of four. There were four focus groups for participants with lived experience (aged 13-18 years) seeking help for themselves. In each of these groups, there were four to five participants, with a total of 17 participants. Among these participants, there were five males, 10 females, one gender diverse individual, and one transgender individual. All participants were reimbursed with an Aus $30 (approximately US $23) online gift voucher for their focus group participation.

Data Collection

Semistructured interview questions were used to guide the focus group discussion based on the capacity to gather in-depth descriptive data. The interview guide was developed by the research team. Example questions included “How likely would you be to recommend KIT to someone else?” and “How easily were you able to find the information you were looking for?”. The questions focused on the content, structure, and design of the chatbot, as well as the design of the chatbot character. The use of a semistructured interview allowed the researchers to explore the research questions, while permitting for flexible responses and any follow-up questions.

The focus groups were conducted by the authors GS and FB, who both have doctoral-level training in psychology and practice as clinical psychologists. All focus groups were conducted online via Zoom, over a period of 2 weeks in October 2020. Focus groups lasted from 61 to 86 minutes, with an average of 75 minutes.

Data Analysis

Qualitative data analysis was conducted using thematic analysis, following the guidelines of Braun and Clarke [29]. The audio recordings for all focus groups were transcribed by a professional transcription service, which were read thoroughly multiple times. The data were coded by the author FB based on basic-level content and then grouped into appropriate categories. The themes were developed to capture the key aspects of participants’ responses, which were regularly reviewed by GS.

Ethics

This project was approved by the Monash University Human Research Ethics Committee (MUHREC ID: 22527). All participants provided either written or audio-recorded verbal consent to participate in the focus group and the audio recording of the focus group, with parental/guardian consent recorded for all participants under the age of 18 years.

Results

The results of the qualitative analysis are presented together for both types of focus groups as the overarching themes were identical (chatbot character and design, content presentation, and flow). Quotes for each of the subthemes are presented in Table 1.

Table 1. Themes, subthemes, and quotes from the focus groups.

Themes and subthemes	Quote number	Data extraction
Key theme 1: Chatbot character and design		
Diverse appeal	1	*To me, KIT felt like non-binary, and I know it's like a ridiculous thing, but I felt seen by this little character drawing, and I think that that never happens, so I really appreciate that, especially because gender diverse folks and eating disorders, there's a really high correlation of the two.* [age 17 years, gender diverse, "help for me" group]
Diverse appeal	2	*I like how it's very gender non-specific. It doesn't scream like either male or female or anything like that, it's just very basic, very good, so that appeals to everyone.* [age 17 years, transgender male, "help for me" group]
Diverse appeal	3	*I also like how KIT doesn't have like really a body as much? I think that's a good thing to avoid, triggers and getting to that minefield.* [age 18 years, male, "help for me" group]
Diverse appeal	4	*I love the fact that KIT is essentially just a head with arms coming out so we've taken away any sense that it could be a body and I think that's really important.* [age 51 years, male, "parent/carer" group]
Diverse appeal	5	*I like its expressions and I love that you've incorporated all those in non-gender way, it's fantastic.* [age 53 years, female, "parent/carer" group]
Visual appeal	6	*I think like helpful, problem-solving, like that second expression with the light-bulb, it's like "let's figure something out here." I was going to say helpful, curious, that kind of thing.* [age 18 years, male, "help for me" group]
Visual appeal	7	*I think he looks really kind of positive and cheerful. I think automatically, all of his expressions look really happy would come across to me.* [age 15 years, female, "help for me" group]
Visual appeal	8	*Ten out of ten! Wholesome, very good looking bot. I love it!* [age 17 years, female, "help for me" group]
Visual appeal	9	*I think it's like kind and simple? Like it's uncomplicated and I think that's a really positive thing because often when you're looking for information, it's really full-on and whereas this just looks like "oh this is going to be simple."* [age 53 years, female, "parent/carer" group]
Visual appeal	10	*It's got a cheery disposition so you know, it probably encourages you to at least investigate what's on offer and if you're feeling down or lost, hopefully you can get something positive out of.* [age 54 years, female, "parent/carer" group]
Key theme 2: Content presentation		
Brevity	11	*I feel like there's a lot of information in there and it's all really good information, but it's all kind of piled onto you, and that can be quite overwhelming for some people.* [age 15 years, female, "help for me" group]
Brevity	12	*I definitely think for a chatbot, you want it to be little text things, not full chunks so I think all the information and the content was really good but if you did split it up into smaller bits, that you could click "more" if you wanted to see more. It would make it seem more like an interactive chat.* [age 15 years, female, "help for me" group]
Brevity	13	*Because it's in big chunks of text, it sometimes gets a bit hard to read and I liked the ones where it had emoji dot points.* [age 17 years, transgender male, "help for me" group]
Brevity	14	*I love the way that the information is simplified and then there's the link that goes to the detailed pages, because I remember when you're in that initial teary "mum phase," when you were going straight to those pages, it was just "Whoa!" whereas this is "oh, go here for the more information" and that gives your brain the time to adjust and be ready so I loved that about it.* [age 53 years, female, "parent/carer" group]
Brevity	15	*I think it's important to have the links to allow people to go and look for more information for when you're at that point and ready to seek more, but I think what everyone's saying about the need for it to be more punchy in the speech bubble is probably the key thing. You must want the facts I guess, but you also need to be able to look further if you want additional information.* [age 46 years, female, "parent/carer" group]

Themes and subthemes	Quote number	Data extraction
Tone	16	*It was worded very factually and formally which is good when you're looking at facts, but I think to make it more approachable for people if you used more informal language but obviously still making it accurate information, I think that would make it seem easier to process for people.* [age 15 years, female, "help for me" group]
Tone	17	*If somebody clicks something like you know the bit where it's like, "seeking help for myself," you're like, "I'm really glad that you've reached out to us." Or if somebody clicks a particular button, that's so valid. Here's some resources.* [age 17 years, gender diverse, "help for me" group]
Tone	18	*Even if you had a line that said at the start, "it is not uncommon to feel alone in this." Because that's probably how support people are feeling, however help is available to reduce your isolation and your fears or whatever like that.* [age 54 years, female, "parent/carer" group]
Key theme 3: Flow		
Navigation	19	*I like the idea of just buttons. I think it's just easier to use. Especially because I feel like it's hard to describe it sometimes – what you're feeling and what you want to know about, so having the buttons just makes it easier because you can just look at what you want.* [age 17 years, male, "help for me" group]
Navigation	20	*I do like the idea of having obviously more things come up as you're reading one because then it feels a lot more flowy, as opposed to seeing something going back. It feels like you're having more of a conversation.* [age 17 years, transgender male, "help for me" group]
Navigation	21	*I think a lot of people going on it might not know specifically what they're looking for, so having to type it all out, you might not even know what you want. So I think the buttons are good because it's all laid out there in front of you instead of looking for anything.* [age 15 years, female, "help for me" group]
Navigation	22	*I think the buttons are okay, but I wonder if they can be ordered differently. I guess if you're going into something like this, you're probably looking for more immediate kind of answers and help so as a carer, you're probably looking at the "what should I look out for?"* [age 46 years, female, "parent/carer" group]
Navigation	23	*I found the navigating really easy and the way that it told you what buttons you'd pushed in the history. That's really good.* [age 53 years, female, "parent/carer" group]
Purpose	24	*I think we're just in such a hurry for information so I was looking for a survey that would tell whether my kid is okay or not. I just want to know if he's okay.* [age 55 years, female, "parent/carer" group]
Purpose	25	*I would be giving them step-by-step instructions. This will empower you when you go to the doctor [with your child].* [age 54 years, female, "parent/carer" group]
Purpose	26	*To have something that people can use 24/7 is really important when they can't access any other help or just to get ideas of where to go.* [age 54 years, female, "parent/carer" group]

Key Theme 1: Chatbot Character and Design

Diverse Appeal

Overall, both groups of participants were satisfied with the appearance of KIT, the character, and reported that it would appeal to a wide audience of users. Favorable comments were made regarding the nongender specific design of KIT (quotes 1, 2, and 5) as participants reported this would appeal to users across the gender spectrum. This was noted to be important given the high correlation of eating disorders and body image concerns among gender diverse people [30].

Both groups of participants found the nonhuman design of KIT to be important as it felt safe and unlikely to trigger body image concerns (quotes 3 and 4). The participants in the parent/carer role provided particularly favorable comments regarding the nonhuman design due to multiple concerns raised that young people would compare their appearance to the character. It was expressed that "the less body parts the better" for the chatbot character. There was some discussion regarding the age appropriateness of KIT's character, as both groups reported that it was slightly more targeted toward youth, but overall did not appear too "childish." However, the parents/carers reported that it was more important for the character to appeal to young people and users rather than solely appealing to older users.

The young people in the "help for me" focus groups commented on the importance of including the pictorial character of KIT within the conversation text so it still seemed like they were conversing with "someone," even though they knew this was a nonhuman chatbot. Interspersing images of KIT, the character, within the lengthy written information increased engagement and useability.

Visual Appeal

Both groups were positive regarding the design and style of KIT's character and described KIT as "wholesome," "engaging," "resourceful," and "nonjudgemental" (quotes 6-10). KIT's expressions were reported by both groups to appeal to young people in particular and were perceived as engaging and approachable. Overall, the two groups of participants approved of the color scheme, with the majority of young people perceiving blues and greens as calming colors. When asked to describe their first impressions of KIT, the opinions of both groups were overwhelmingly positive, including "knowledgeable," "inviting," "approachable," "kind," "fun," "professional," and "cheerful." These impressions centered around the idea of KIT as a character who could provide resources and high-quality information, while also appearing friendly and engaging (quote 6).

In addition to the visual design of the character, parent/carer and young participants commented upon the design of the webchat. Of importance to the parents/carers was the design, which was perceived as simple and thus promoting useability, as compared to other mental health resources that can appear "complicated" and "overwhelming." The design of the webchat page was described as "clean and modern," which was reported to be appealing and accessible for both groups of participants.

Key Theme 2: Content Presentation

Brevity

The majority of participants in the two groups regarded the content of the webchat as positive and reported that this would be best presented in smaller sections (than was included in the prototype) to avoid overwhelming the user (quote 11). The preference for brevity was noted by both groups. It was expressed that individuals seeking this information may be feeling stressed and scared, and experiencing a sense of urgency; therefore, "short and sharp" content was reported to be more supportive.

Both groups of participants liked the use of emojis as dot points to enhance readability of the text (quote 13). In addition, the ideas of staggered messages, a "click for more" button and putting the content into multiple message bubbles (quote 12) were also raised to optimize the presentation of the written information within the chatbot. Some participants related this to receiving a long message from a friend and preferring multiple shorter messages instead. The young people and parents/carers reported that breaking up the content into "chunks" would feel more "conversational and interactive." Participants from both groups appreciated the strategy of providing some information with additional resources available via external weblinks (quotes 14 and 15). This format was reported to be more inviting and accessible so users could explore the content at their own pace.

Tone

The interactivity of KIT was discussed in relation to the tone. Most participants from both groups reported that in addition to the factual information KIT presented, KIT could include supportive statements (eg, "I'm really glad that you've reached out to us" [quote 17]), which would provide a more conversational tone as well as normalize any concerns of the user (quote 17). Amongst the younger people in the "seeking help for myself" group, these participants preferred a more casual tone of conversation, while acknowledging it was necessary to present the information accurately (quote 16).

The parent/carer participants reported comments regarding the guided nature of the content, so they could feel reassured by the information, given their own experiences of feeling isolated, scared, and lost, as carers of young people with eating disorders (quote 18). Parents/carers reported that it was important not to "soften" the information, and alternatively present factual and realistic guidance, given the severity and risk of death associated with eating disorders [4]. All participants from both groups agreed that it was important for the content to be presented as factual and accessible, while balancing the supportive tone required.

Key Theme 3: Flow

Navigation

The majority of participants in both groups reported that the button-based navigation was sufficient in navigating the chatbot rather than free text. The younger participants reported this was useful for people who might not know specifically what information or support they require (quotes 19 and 21). The buttons appealed to broader accessibility, with parent/carer and younger participants citing that the buttons would be beneficial for those with language difficulties. The buttons were reported to cover the relevant information, and the navigation was described as simple and straightforward to use (quote 23).

Some participants in both groups expressed that a free text option would be beneficial for more specific questions, but also described frustration at the current level of artificial intelligence technology available in other chatbots whose natural language processing was not sufficiently capable of understanding their typed queries. Generally, the parents/carers expressed a preference for having both button-directed navigation as well as the option to type in specific questions. These participants expressed that individuals using the "help for another" pathway would likely already be concerned about their loved one and would by this point have specific questions to which they would be seeking answers.

Regarding the flow of the chatbot conversation, some parents/carers and younger people suggested having "categories" where a user could flow through these suggested themes or having "next" buttons for relevant sections. Improving the flow of the conversation was considered important to enhance the interactivity of the experience. Some young people reported getting "decision fatigue" when "overwhelmed" with multiple buttons, and reported that merging the sections into categories would be helpful. The main feedback from parents/carers was to change the order of how information was presented in the buttons, as they reported that the immediate questions of "what do I look out for?" and "how can I help my child?" were the most important (quote 22). Both groups of participants appreciated seeing the history of the chat and the ability to scroll back and reread the conversation.

Purpose

One point of difference between the two groups of focus group participants was that the parent/carer group discussed the purpose and capabilities of the chatbot. This included queries around the proposed audience of the chatbot and how it might differ from other available resources. Some participants reported conversing with the chatbot with a specific question in mind, such as “Should I be worried about my child?” or “Where can I find a psychologist?” (quote 24). The researchers provided feedback regarding the ethical and practical limitations of the chatbot, with the purpose focused on providing psychoeducation and “in the moment” coping skills. It was apparent that the younger users were more familiar with the purpose and limitations of such technology, whereas the parents/carers presented different expectations regarding the specificity of their requests (quote 25). Both groups of participants praised the 24/7 availability of the chatbot as a major strength, particularly when other body image and eating disorder–focused supports are not available 24/7 (quote 26). In summary, the parents/carers reported that the chatbot would likely be a useful resource, and acknowledging the capabilities and limitations of the technology from the start of the conversation with KIT would be helpful to manage user expectations.

Discussion

Principal Findings

The aim of this study was to explore the preliminary acceptability and feasibility of a world-first chatbot designed to support people experiencing body image and eating concerns, and parents/carers. The qualitative results from the focus groups illustrated three main themes as follows: chatbot character and design, content presentation, and flow. Across these themes, both people seeking help for themselves and parents/carers generally provided positive feedback. The chatbot was praised for its clean design, ease of navigation, and engaging character. Both the groups approved of the gender nonconforming and nonhuman design of KIT, which was believed to improve the accessibility of the chatbot and help diverse users feel safe. Most participants approved of KIT’s facial expressions and found the character approachable, resourceful, and calming.

Most participants reported the content was appropriate, accessible, and helpful. To improve the chatbot, both audiences recommended presenting the amount of information differently, with a preference for “short and sharp” content. This preference for brevity was balanced by implementing a supportive tone and using opportunities to normalize concerns as well as provide information.

The participants seeking help for themselves preferred the navigational style using buttons, whereas the parents/carers requested the option to type specific queries as well as be guided through the content via buttons. This difference in preferences likely reflected the different mindsets these two audiences expressed. The young people preferred to be guided with the options laid out, whereas parents/carers often had specific queries and concerns. Overall, the participants expressed positive feedback regarding the body image chatbot, and specific recommendations were implemented in the development of KIT as per iterative agile methodology [31].

Comparison With Prior Work

To our knowledge, this is the first chatbot designed to specifically target body image and eating concerns, with the majority of other digital mental health interventions focused on mood and anxiety disorders [24]. In addition, reports specifically addressing the development and design of chatbots in the field of mental health are rare [21,32]. However, our study found that users valued the chatbot’s useability, quality of content, and engagement, which was similar to previous research [33].

One area of divergence from the existing literature was related to the subtheme of brevity. A systematic review of mental health chatbots illustrated that most participants perceived conversations as too short or shallow [33]. However, in this study, participants reported a preference for shorter information. It is possible this difference reflects KIT’s strong focus on psychoeducational information, but nevertheless highlights the necessity of providing users with appropriate length responses for their needs.

Within the literature, the empathy of nonhuman agents has received much research attention, with evidence indicating the importance of the development of therapeutic bonds [20,34,35]. While there were recommendations to increase the interactivity of KIT within conversations, the use of a pictorial character within conversation text appeared to be a useful “shortcut” for building rapport with users.

Future research with larger sample sizes is required regarding the use and efficacy of KIT, which will provide user feedback and allow for further refinements of KIT’s content. After such evaluations, KIT can be more easily compared with other pre-existing and well-established mental health chatbots such as Woebot and Wysa [13,34,36]. With KIT’s focus on delivery of coping skills, particularly based on CBT, KIT is in line with chatbots like Woebot and Wysa, but has the unique focus of body image and eating issues compared to anxiety and depression [24].

Strengths and Limitations

The main strength of this study was the diverse recruitment of participants for the focus groups, which strongly benefitted the co-design process of KIT. Both the parent/carer and young people groups included individuals from all regions of Australia owing to the online format. In the young people group, care was taken to include participants across the gender spectrum, including transgender and gender diverse individuals. This process included ensuring there was a mix of genders in each focus group. The majority of participants were female, which is reflective of the higher prevalence of eating disorders among women and girls. The study was limited by a smaller sample size. Furthermore, the sample was diverse in all but one demographic feature, which was ethnicity, with most participants identifying as Caucasian.

Given that participants were recruited through online advertising, bias is inherent in those who can access and respond to research advertisements. In addition, a small percentage of participants had previously been involved in preliminary online co-design forums for KIT, which may have influenced their opinions and feedback in the subsequent focus groups. Social desirability may have also influenced the feedback provided by participants, as the authors were conducting the focus groups. However, the authors strongly encouraged honest responses, and a range of positive, negative, and neutral comments were made by participants.

Conclusions

Our study findings provide preliminary support for the acceptability and feasibility of KIT, the body image chatbot. The results showed overall positive feedback regarding the content, structure, and design of the body image chatbot, from both groups of participants in the "seeking help for myself" and "seeking help for another" pathways. The participants' recommendations and feedback were used to refine KIT prior to public launch in order to improve the layout, content, and navigation of the chatbot. If proven effective, KIT would likely provide a valuable resource of psychoeducation and coping skills by meeting young people where they spend their time (online) and those who might not feel ready to follow traditional formats of seeking help. KIT may assist in filling the gaps in service delivery in terms of both prevention and intervention by supporting people and those concerned about them, with appealing and accessible evidence-based psychoeducation and skills.

Acknowledgments

The authors would like to thank the focus group participants for their valuable participation and for sharing their lived experience, and Dr Sally Bradford for her invaluable support in the earliest stages of this project. The authors would also like to thank Dr Sarah Rotstein and Marcia Scott for their creative support. The development of KIT was a collaboration with the Butterfly Foundation (Melissa Wilton, Belinda Kerslake, Juliette Thomson, Alex Cowen, and Michaela Del Casale), chatbot production company Proxima (Sebastian Pedavoli, Dan Nolan, and Roshni Roy), and graphic design team Yoke (Sarah Lawrey, Ryan Nadzielski, and Krysa Allison). GS was the recipient of an AMP Tomorrow Maker Grant to fund this project and holds a National Health and Medical Research Council (NHMRC) Early Career Fellowship (GNT1146033). SLR holds a Senior NHMRC Fellowship (GNT1154651).

Conflicts of Interest

None declared.

References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
2. Trautmann S, Rehm J, Wittchen H. The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? EMBO Rep 2016 Sep 04;17(9):1245-1249 [FREE Full text] [doi: 10.15252/embr.201642951] [Medline: 27491723]
3. Butterfly Report: Paying the Price-The impact of eating disorders in Australia. Deloitte. 2012. URL: https://www2.deloitte.com/au/en/pages/economics/articles/butterfly-report-paying-price-eating-disorders.html [accessed 2021-08-04]
4. Smink FRE, van Hoeken D, Hoek HW. Epidemiology of eating disorders: incidence, prevalence and mortality rates. Curr Psychiatry Rep 2012 Aug;14(4):406-414 [FREE Full text] [doi: 10.1007/s11920-012-0282-y] [Medline: 22644309]
5. Galmiche M, Déchelotte P, Lambert G, Tavolacci M. Prevalence of eating disorders over the 2000-2018 period: a systematic literature review. Am J Clin Nutr 2019 May 01;109(5):1402-1413. [doi: 10.1093/ajcn/nqy342] [Medline: 31051507]
6. Butterfly Report - Investing in Need: Cost Effective Interventions for Eating Disorders. Butterfly Foundation. 2015. URL: https://tinyurl.com/u5t689sp [accessed 2021-04-08]
7. Insights in Body Esteem: A survey of Australians' experience of body image and its impact on day to day life. Butterfly Research Institute. 2018. URL: https://butterfly.org.au/wp-content/uploads/2020/05/Insights-into-Body-Esteem-Report-PDF2.pdf [accessed 2021-04-08]
8. Tiggemann M, Anderberg I, Brown Z. Uploading your best self: Selfie editing and body dissatisfaction. Body Image 2020 Jun;33:175-182. [doi: 10.1016/j.bodyim.2020.03.002] [Medline: 32224447]
9. Fardouly J, Diedrichs PC, Vartanian LR, Halliwell E. Social comparisons on social media: the impact of Facebook on young women's body image concerns and mood. Body Image 2015 Mar;13:38-45. [doi: 10.1016/j.bodyim.2014.12.002] [Medline: 25615425]
10. Tankovska H. Number of global social network users 2017-2025. Statista. URL: https://tinyurl.com/2hepr955 [accessed 2021-04-08]
11. Heruc G, Hurst K, Casey A, Fleming K, Freeman J, Fursland A, et al. ANZAED eating disorder treatment principles and general clinical practice and training standards. J Eat Disord 2020 Nov 10;8(1):63 [FREE Full text] [doi: 10.1186/s40337-020-00341-0] [Medline: 33292546]

12. Atwood ME, Friedman A. A systematic review of enhanced cognitive behavioral therapy (CBT-E) for eating disorders. Int J Eat Disord 2020 Mar 16;53(3):311-330. [doi: 10.1002/eat.23206] [Medline: 31840285]
13. Jones RB, Thapar A, Rice F, Mars B, Agha SS, Smith D, et al. A Digital Intervention for Adolescent Depression (MoodHwb): Mixed Methods Feasibility Evaluation. JMIR Ment Health 2020 Jul 17;7(7):e14536 [FREE Full text] [doi: 10.2196/14536] [Medline: 32384053]
14. Fernando AN, Sharp G. Genital Self-Image in Adolescent Girls: The Effectiveness of a Brief Educational Video. Body Image 2020 Dec;35:75-83. [doi: 10.1016/j.bodyim.2020.08.007] [Medline: 33011539]
15. Alleva JM, Sheeran P, Webb TL, Martijn C, Miles E. A Meta-Analytic Review of Stand-Alone Interventions to Improve Body Image. PLoS One 2015 Sep 29;10(9):e0139177 [FREE Full text] [doi: 10.1371/journal.pone.0139177] [Medline: 26418470]
16. Hurst K, Heruc G, Thornton C, Freeman J, Fursland A, Knight R, et al. ANZAED practice and training standards for mental health professionals providing eating disorder treatment. J Eat Disord 2020 Nov 02;8(1):58 [FREE Full text] [doi: 10.1186/s40337-020-00333-0] [Medline: 33292542]
17. Wade T, George WM, Atkinson M. A randomized controlled trial of brief interventions for body dissatisfaction. J Consult Clin Psychol 2009 Oct;77(5):845-854. [doi: 10.1037/a0016879] [Medline: 19803565]
18. Fuller-Tyszkiewicz M, Richardson B, Lewis V, Linardon J, Mills J, Juknaitis K, et al. A randomized trial exploring mindfulness and gratitude exercises as eHealth-based micro-interventions for improving body satisfaction. Computers in Human Behavior 2019 Jun;95:58-65. [doi: 10.1016/j.chb.2019.01.028]
19. 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]
20. 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]
21. Høiland CG, Følstad A, Karahasanovic A. Hi, Can I Help? Exploring How to Design a Mental Health Chatbot for Youths. Human Technology 2020 Aug 31;16(2):139-169. [doi: 10.17011/ht/urn.202008245640]
22. Vaidyam AN, Linggonegoro D, Torous J. Changes to the Psychiatric Chatbot Landscape: A Systematic Review of Conversational Agents in Serious Mental Illness: Changements du paysage psychiatrique des chatbots: une revue systématique des agents conversationnels dans la maladie mentale sérieuse. Can J Psychiatry 2021 Apr;66(4):339-348. [doi: 10.1177/0706743720966429] [Medline: 33063526]
23. Abd-Alrazaq A, Rababeh A, Alajlani M, Bewick B, Househ M. Effectiveness and Safety of Using Chatbots to Improve Mental Health: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
24. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: A scoping review. Int J Med Inform 2019 Dec;132:103978. [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
25. Linardon J, Wade TD, de la Piedad Garcia X, Brennan L. The efficacy of cognitive-behavioral therapy for eating disorders: A systematic review and meta-analysis. J Consult Clin Psychol 2017 Nov;85(11):1080-1094. [doi: 10.1037/ccp0000245] [Medline: 29083223]
26. Linardon J, Gleeson J, Yap K, Murphy K, Brennan L. Meta-analysis of the effects of third-wave behavioural interventions on disordered eating and body image concerns: implications for eating disorder prevention. Cogn Behav Ther 2019 Jan;48(1):15-38. [doi: 10.1080/16506073.2018.1517389] [Medline: 30307377]
27. Fast E, Chen B, Mendelsohn J, Bassen J, Bernstein MS. Iris: A conversational agent for complex tasks. In: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, QC, Canada p. 1-12. [doi: 10.1145/3173574.3174047]
28. Are they old enough: how to navigate key online milestones for your child. eSafety Commissioner. URL: https://www.esafety.gov.au/parents/skills-advice/are-they-old-enough [accessed 2021-04-08]
29. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
30. Strübel J, Sabik NJ, Tylka TL. Body image and depressive symptoms among transgender and cisgender adults: Examining a model integrating the tripartite influence model and objectification theory. Body Image 2020 Dec;35:53-62. [doi: 10.1016/j.bodyim.2020.08.004] [Medline: 32846389]
31. Kannan V, Basit MA, Youngblood JE, Bryson TD, Toomay SM, Fish JS, et al. Agile co-development for clinical adoption and adaptation of innovative technologies. 2017 Presented at: 2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT); November 6-8, 2017; Bethesda, MD p. 56-59. [doi: 10.1109/hic.2017.8227583]
32. Park S, Choi J, Lee S, Oh C, Kim C, La S, et al. Designing a Chatbot for a Brief Motivational Interview on Stress Management: Qualitative Case Study. J Med Internet Res 2019 Apr 16;21(4):e12231 [FREE Full text] [doi: 10.2196/12231] [Medline: 30990463]

33. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review. J Med Internet Res 2021 Jan 13;23(1):e17828 [FREE Full text] [doi: 10.2196/17828] [Medline: 33439133]
34. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
35. Lucas GM, Gratch J, King A, Morency L. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 2014 Aug;37:94-100. [doi: 10.1016/j.chb.2014.04.043]
36. 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]

Abbreviations

ACT: acceptance commitment therapy
CBT: cognitive behavioral therapy
CBT-E: enhanced cognitive behavioral therapy

Edited by R Kukafka; submitted 12.02.21; peer-reviewed by R Pine, D Levine, A Fernandes; comments to author 26.03.21; revised version received 08.04.21; accepted 24.05.21; published 16.06.21.

Please cite as:
Beilharz F, Sukunesan S, Rossell SL, Kulkarni J, Sharp G
Development of a Positive Body Image Chatbot (KIT) With Young People and Parents/Carers: Qualitative Focus Group Study
J Med Internet Res 2021;23(6):e27807
URL: https://www.jmir.org/2021/6/e27807
doi:10.2196/27807
PMID:34132644

Original Paper

Weight Loss and Usage of an Online Commercial Weight Loss Program (the CSIRO Total Wellbeing Diet Online) Delivered in an Everyday Context: Five-Year Evaluation in a Community Cohort

Gilly A Hendrie[1], PhD; Danielle L Baird[1], BSc; Emily Brindal[1], PhD; Gemma Williams[1], BSc; Jennie Brand-Miller[2], PhD; Beverly Muhlhausler[1], PhD

[1]Nutrition and Health Program, Commonwealth Scientific and Industrial Research Organisation, Adelaide, Australia

[2]Charles Perkins Centre and School of Life and Environmental Sciences, University of Sydney, Sydney, Australia

Corresponding Author:
Gilly A Hendrie, PhD
Nutrition and Health Program
Commonwealth Scientific and Industrial Research Organisation
PO Box 10041
Adelaide, 5000
Australia
Phone: 61 8 8305 0662
Email: GILLY.HENDRIE@CSIRO.AU

Abstract

Background: Obesity is a global public health challenge, and there is a need for more evidence-based self-management programs that support longer-term, sustained weight loss.

Objective: This study used data from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Total Wellbeing Diet Online program to determine the reach and weight loss results over its first 5 years.

Methods: Participants were adults who joined the commercial weight loss program of their own volition between October 2014 and September 2019 (N=61,164). Information collected included year of birth, sex, height, weight, and usage data (eg, entries into the food diary, views of the menu, and program content). Weight loss and percentage of starting body weight lost were calculated. Members were divided into 2 groups for analysis: “stayers” were members who signed up for at least 12 weeks of the program and recorded a weight entry at baseline and at the end of the program, while “starters” began the program but did not record a weight after 12 weeks. Descriptive statistics and multiple linear regression were used to describe weight loss and determine the member and program characteristics associated with weight loss.

Results: Data were available from 59,686 members for analysis. Members were predominately female (48,979/59,686, 82.06%) with an average age of 50 years (SD 12.6). The average starting weight was 90.2 kg (SD 19.7), and over half of all members (34,195/59,688, 57.29%) were classified as obese. At week 12, 94.56% (56,438/59,686) of the members had a paid program membership, which decreased to 41.48% (24,756/59,686) at 24 weeks. At week 12, 52.03% (29,115/55,958) of the remaining members were actively using the platform, and by week 24, 26.59% (14,880/55,958) were using the platform. The average weight loss for all members was 2.8 kg or 3.1% of their starting body weight. Stayers lost 4.9 kg (5.3% of starting body weight) compared to starters, who lost 1.6 kg (1.7% of starting body weight). Almost half (11,082/22,658, 48.91%) the members who stayed on the program lost 5% or more of their starting body weight, and 15.48% (3507/22,658) achieved a weight loss of 10% or more. Of the members who were classified as class 1 obese when they joined the program, 41.39% (3065/7405) who stayed on the program were no longer classified as obese at the end, and across all categories of obesity, 24% (3180/13,319) were no longer classified as obese at the end of the program. Based on multiple linear regression, platform usage was the strongest predictor of weight loss (β=.263; P<.001), with higher usage associated with greater weight loss.

Conclusions: This comprehensive evaluation of a commercial, online weight loss program showed that it was effective for weight loss, particularly for members who finished the program and were active in using the platform and tools provided. If the results demonstrated here can be achieved at an even greater scale, the potential social and economic benefits will be extremely significant.

(J Med Internet Res 2021;23(6):e20981) doi:10.2196/20981

KEYWORDS

obesity; obesity management; weight loss; internet-based intervention

Introduction

Background

Obesity is a global public health challenge, with significant social and economic impacts. In Australia, two-thirds of adults are classified as overweight or obese [1], and this is expected to increase to more than three-quarters of the adult population by 2030 [2]. The costs to the Australian economy attributable to overweight status and obesity were estimated to be A $8.6 billion between 2011 and 2012 [1]. There is growing recognition that nutrition plays a crucial role in the etiology of chronic diseases and that chronic diseases such as type 2 diabetes can also be reversed by diet and lifestyle interventions [3]. With most adults struggling to control their weight, there is a need for more evidence-based self-management programs that support healthy diet and lifestyle patterns to promote longer-term, sustained weight loss [4].

In 2020, commercial weight loss services in Australia were estimated to be worth A $452 million [5]. Although there are several programs with a variety of dietary patterns, multiple literature reviews have concluded that weight loss results across programs are similar and that the vast majority produce only small to moderate effects which are not maintained in the longer term [6-8]. Reviews of branded and commercial programs have been primarily based on randomized controlled trials (RCTs) and suggest individuals who complete these programs can achieve significant weight loss [8]. However, in evaluations undertaken through longitudinal follow-up of participants, only a small proportion of the original number of individuals are included [9,10]; therefore, the findings are unlikely to be representative of the full participant population.

Web-Based Delivery of Weight Loss Programs

In an attempt to increase the accessibility and success of commercial weight loss programs, a number of web-based programs have been developed. One early meta-analysis suggested that web-based interventions achieved similar weight loss to controls but that those with enhanced features could achieve greater weight loss than those providing education alone [11]. Some but not all more recent reviews suggest that web-based programs can be more successful than alternate delivery approaches, but the effect sizes are small, and the heterogeneity in study designs makes it difficult to determine the key elements driving weight loss [12-15].

Programs delivered online are growing in popularity and account for 17% of the market share for the diet industry [5]. Commercial programs that have traditionally relied on face-to-face group sessions have now shifted their focus to primarily digital delivery. An early review suggested that self-directed, online commercial programs were suboptimal [16], but since then, more advanced features have improved the offerings in the market, with associated RCTs suggesting moderate effects for weight loss. For example, one RCT that included 309 people on the 12-week Biggest Loser Club program reported weight losses of 2.0-3.2 kg compared to 0.5 kg in a wait-listed control group [17]. More recently, an evaluation of participants receiving the Weight Watchers Online program reported a 2.7 kg weight loss relative to a 1.3 kg loss in a control group receiving only a newsletter [18]. However, in a US study, participants on an eDiets program lost less weight over a 1-year period than did those receiving a comprehensive information manual (1.1% vs 4.0% of starting body weight, respectively). It should be noted that this evaluation was small and included only 47 participants in total [19]. Thus, based on the data from existing RCTs, it appears that online commercial weight loss programs can be successful. However, less has been published about how these programs may work in everyday contexts [16] where attrition is likely to be significant, especially when users are self-directed and incurring possible program costs.

In late 2014, the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Digital Wellness launched a commercial, online version of the CSIRO Total Wellbeing Diet. The dietary components were developed through clinical trials [20-23] and were initially translated into a series of popular books [24], estimated to have delivered weight loss benefits to 290,700 Australians, with an average weight loss of 5.7 kg [25]. The online format delivered the same program as the books did but through a digital platform, allowing for several enhancements, including personalized eating plans, customized weekly meal plans, food and exercise diaries, the ability to record and see progress of weight loss, a member forum, and supportive correspondence via email. Our study analyzed the data from the CSIRO Total Wellbeing Diet Online program in the first 5 years after it launched to determine the program’s reach and weight loss results over this period and to investigate the key determinants of weight loss success. Specifically, we aimed to determine average weight loss and its relationship to the duration of membership, features of the platform that members used most, and the member and usage characteristics that were associated with greater weight loss.

Methods

Study Design and Participants

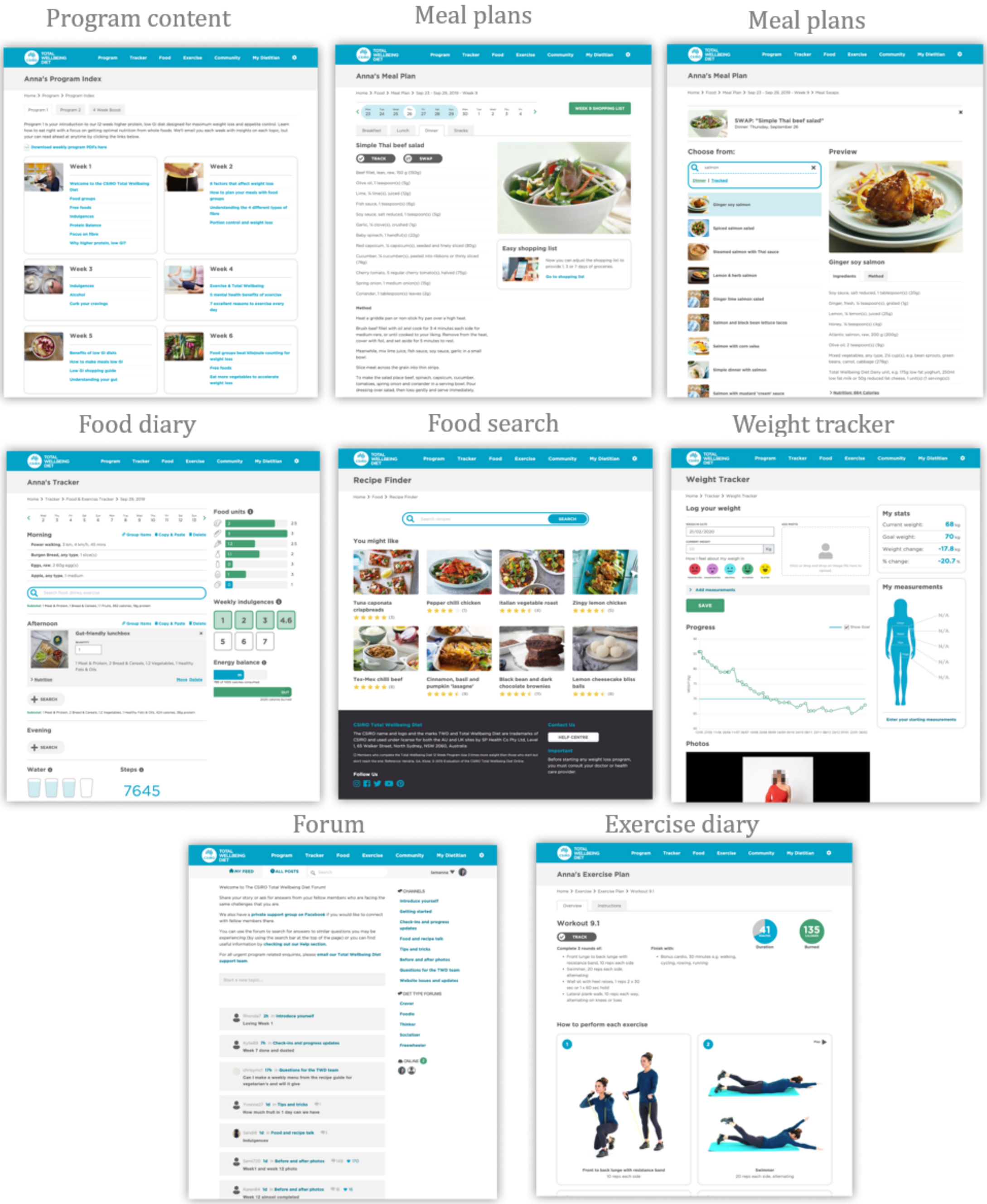
Participants were individuals aged ≥18 years who joined the CSIRO Total Wellbeing Diet Online of their own volition between October 2014 and September 2019 (N=61,164; referred to as “members”). In the registration and setup processes, individuals younger than 18 years or who had a BMI that placed them in the underweight category (BMI <18.5 kg/m^2) were automatically excluded. The participants who were removed from this analysis were as follows: pseudomembers (ie, platform testers or affiliated staff), those whose membership was paid for by their employer because their motivations might have been different to those who signed up and paid for their membership voluntarily, members living outside of Australia because the context in which they were following the program was different and the menu plans were not optimized for seasonal or local produce, and members without a paid subscription. These combined exclusions resulted in 1478 records being removed (1478/61,164; 2.4% of the available data).

Intervention

The CSIRO Total Wellbeing Diet Online [26] is a 12-week higher-protein, lower–glycemic index, commercial weight loss program managed by Digital Wellness and available to individuals at a cost of A $199 for the first 12 weeks. The user registration process collects information on year of birth, sex, physical activity levels, and weight loss goals to tailor eating and exercise plans. Individuals are assigned to 1 of 3 kilojoule bands depending on their starting weight (6000, 7000, or 8000 kJ/day). The diet is structured around 3 meals (breakfast, lunch, and dinner) and 2 snacks each day through use of a food group system where portions of food are presented as standard units for each food group. There are 7 food groups: fruit, vegetables, meat and alternatives, breads and cereals, dairy, healthy fats and oils, and indulgences. Meals are designed around a template of standard units, which ensures daily allowances of food groups are met and provides optimal nutrition and energy to promote weight loss. Daily and weekly meal plans can be customized by swapping meals using the recipe database.

The online platform is a fully responsive web app with an interface that is optimized for viewing on a desktop and on mobile devices. The platform features are designed to support individuals throughout the program (Figure 1). Program content provides general information, nutrition advice, weekly tutorials available for viewing at any time, and some content sent out to members in a weekly email. The meal plans provide individuals with meals and snacks planned for the day or week ahead. The food diary can be used to log meals and snacks consumed, either by entering prepopulated recipes from the meal plan or recipe database, or by entering individual foods from a comprehensive food database. The food search function allows members to search through the food database for individual foods or recipes and view information about their composition and food units. The food tracker tallies the food units and total energy consumed over the day. The exercise diary is where completed exercises can be logged or where activities chosen from a database of different exercise types and intensities can be recorded. Members receive weekly emails on their nominated weigh-in day to remind them to weigh themselves and record their weight into the weight tracker. Progress data are presented in a graph and table form. The platform also has a forum for members to share their stories or discuss relevant issues with other members.

Figure 1. Screenshots of the features of the CSIRO Total Wellbeing Diet Online platform.



Data Collection and Study Outcomes

Digital Wellness manages the collection and storage of data, including registration details and all user activity and interactions with the platform. In addition to information collected at registration, individuals' body measurements, such as height, weight, and waist circumference, are also collected. Other data collected included program details (eg, paid membership duration) and platform usage data, including entries into the food diary, views of the menu plans, views of exercise plans, views of program content information, forum visits, searches of the food database, and weight entries. These data were provided to the research team in a deidentified format with each individual member assigned a unique identifier. As part of registration, participants agreed to their data being used for research purposes; therefore, no direct participant consent was sought. Ethics approval to conduct this research was received from the CSIRO Health and Medical Human Research Ethics Committee (approval #2019_090_LR).

Weight loss was calculated as the difference in kilograms between the last and first weight entered into the platform, with a larger number representing greater weight loss. Weight loss was also calculated as a percentage of starting body weight and was categorized into 4 groups: weight gain (greater than 0% difference), a weight loss of over 0% and less than 5% of starting body weight, 5% to less than 10% of starting body weight, and greater than or equal to 10% of starting body weight lost.

The date of program setup and total duration of paid membership were used to calculate the number of participants remaining in the program within the first 24 weeks from registration. Once paid membership duration lapsed, a member was considered to have left the program (dropped out of the program). Platform usage attrition within the first 24 weeks from registration was also determined by calculating the difference in days between the date of last user activity and the date of the program setup.

Platform usage was considered to be any logged activity and was assessed as usage of each of the 7 features and as overall usage (all features combined). Platform usage was described in 3 ways: “total days active” referred to the number of unique days a member used the platform irrespective of whether usage occurred multiple times per day, “percent active days” was calculated as total days active divided by the number of days of membership expressed as a percentage, and “activity per day” was calculated as the total user activity divided by the number of days of membership. Levels of platform usage were then calculated by creating quintiles (5 groups) based on activity per day, where quintile 1 represented the lowest platform activity and quintile 5 the highest platform activity. There was a small group of participants with very high use, and thus the creation of quintiles helped to manage this skewness.

Statistical Analysis

All data were inspected for invalid records through a systematic and previously used data cleaning process based on erroneous height (less than 1 m or greater than 3 m) or weight values (less than 13 kg or greater than 250 kg), and extreme BMI values (less than 13 kg/m^2 or greater than 97 kg/m^2). Members were removed from analysis if a total weight loss could not be calculated (n=1451) due to no starting weight reported or only a single weight entry being made. Members were also removed from analysis if their weight value was the contributing factor to an invalid BMI (n=14) or if age at registration was calculated from entered data as less than 18 years or above 100 years of age (n=7). In addition, there were 6 pseudomembers removed from analysis. After these exclusions, 59,686 members were included in this analysis (59,686/61,164, 97.58% of all members).

Members were divided into 2 groups for analysis: “stayers” were defined as members who signed up to at least 12 weeks of the program and entered their weight into the platform at baseline and at the end of the program. This was calculated as members with a paid membership equal to or greater than 84 days (12-week program duration) and for whom days between their first and last entered weight were equal to or greater than 77 days (plus or minus 1 week for first or last weigh-in). There were 22,658 members who were categorized as stayers, and the average time between first and last weigh-in was 268 days. Starters were defined as members who started the program and had some level of engagement with the platform but did not enter a weight at the end of the 12-week program. By definition, starters had a shorter duration of paid membership or a last-entered weight that was before 77 days after baseline. There were 37,028 members who were categorized as starters with an average of 23 days between the first and last weigh-in.

The average (and SD) weight loss and percentage weight loss were calculated at a group level for all members and by subgroups of interest. These subgroups included those of sex, age (19-30 years, 31-50 years, 51-70 years, and 70 years and over), starting BMI category (normal weight, overweight, obese class 1, obese class 2, and obese class 3), and socioeconomic status according to quintiles of Socio-Economic Indexes for Areas (SEIFA), where a lower quintile represents a greater disadvantage [27]. Differences in weight loss, percentage weight loss, and platform usage were examined between subgroups of members, with significance being tested by 1-way analysis of variance (ANOVA).

Multiple linear regression was used to assess which member characteristics (sex, age, SEIFA quintile, starting BMI) or program characteristics (membership length, activity per day) were the strongest predictors of weight loss. The primary analysis used the last weight entered into the platform (ie, the last observation carried forward). The models were run to predict total weight loss on the program (using total weight loss in kilograms and as a percentage of starting body weight). Member and program characteristics (predictors) were added to the model simultaneously. A secondary sensitivity analysis was conducted to determine the robustness of the primary results with baseline weight carried forward being used when participants did not have a weight value available in the platform at 12 weeks or beyond. The regression models were run for all members and by subgroup for starters and stayers separately. The significant predictors of weight loss were similar to those used for total weight loss and percentage of body weight lost; therefore, only the regression results for all members predicting percentage of body weight lost are presented. Statistical analyses were performed using SPSS Statistics 25 (IBM Corp).

Results

Characteristics of Members

Members who signed up to the program were predominately female (48,979/59,686; 82.06%) with an average age of 50 years (SD 12.6). Overall, 40.16% (23,969/59,686) of members were aged between 31-50 years, and 48.70% (29,068/59,686) were between 51-70 years (Table 1). The average starting weight of members when they joined the program was 90.2 kg (SD 19.7), the average BMI was 32.2 (SD 6.3; Table S1, Multimedia Appendix 1), and over half of all members were classified as obese (34,195/59,686, 57.29%). More specifically, 30.75% (18,353/59,686) of members were classified as class 1 obese, 16.11% (9613/59,686) as class 2, and 10.44% (6229/59,686) as class 3, the highest risk group (Table 1). Men were heavier than women (104.9 kg vs 87.0 kg, respectively), but the starting BMI of men and women was similar (32.8 vs 32.0, respectively; Table S1, Multimedia Appendix 1).

Members resided in all states of Australia (data not shown). One-quarter (15,005/59,686, 25.14%) of members lived in areas classified in the lowest 2 quintiles of disadvantage (most disadvantaged), and 32.84% (19,598/59,686) lived in areas classified as the least disadvantaged (Table 1). Members living in the most disadvantaged areas were heavier (94.1 kg vs 87.7 kg) and had a higher BMI (33.9 vs 31.1) compared to those in the least disadvantaged areas (Table S1, Multimedia Appendix 1).

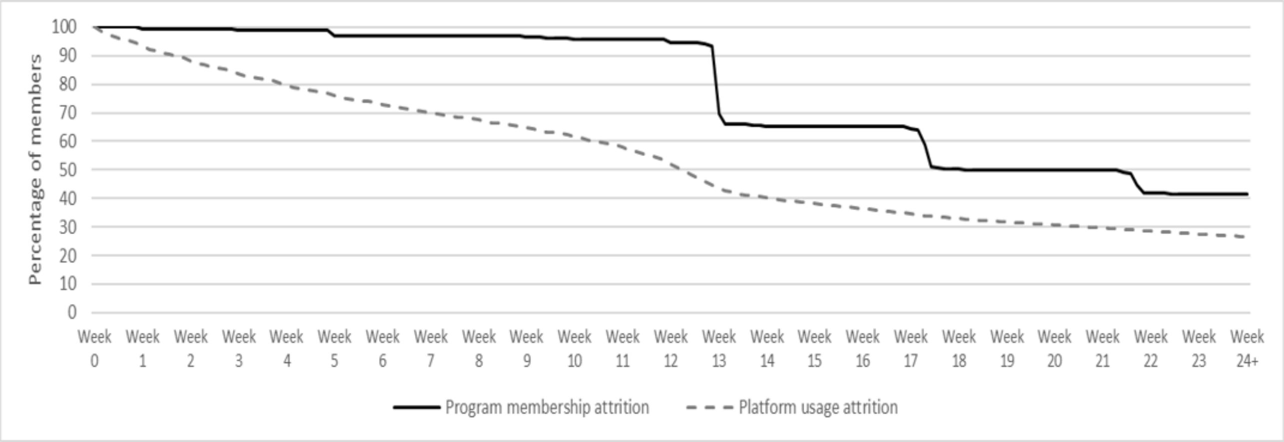
Table 1. Demographic characteristics of starters (n=37,028), stayers (n=22,658), and all members (N=59,686) at the time of joining the CSIRO Total Wellbeing Diet Online.

Member characteristics	Starters, n (%)	Stayers, n (%)	All members, n (%)
Sex			
Male	6204 (16.75)	4503 (19.87)	10,707 (17.94)
Female	30,824 (83.25)	18,155 (80.13)	48,979 (82.06)
Age (years)			
18-30	2312 (6.24)	1004 (4.43)	3316 (5.56)
31-50	15,766 (42.58)	8203 (36.20)	23,969 (40.16)
51-70	17,108 (46.20)	11,960 (52.78)	29,068 (48.70)
>71	1842 (4.97)	1491 (6.58)	3333 (5.58)
Starting BMI category			
Normal weight	3771 (10.19)	1473 (6.51)	5244 (8.79)
Overweight	11,171 (30.19)	7786 (34.39)	18,957 (31.78)
Obese	20,859 (56.37)	13,336 (58.90)	34,195 (57.33)
Class 1	10,937 (29.55)	7416 (32.75)	18,353 (30.77)
Class 2	5956 (16.09)	3657 (16.15)	9613 (16.12)
Class 3	3966 (10.72)	2263 (10.00)	6229 (10.44)
Missing/invalid	1205 (3.26)	46 (0.20)	1251 (2.10)
Socioeconomic status			
1 (lowest)	3665 (9.90)	2322 (10.25)	5987 (10.03)
2	5579 (15.07)	3439 (15.18)	9018 (15.11)
3	6971 (18.83)	4446 (19.62)	11,417 (19.13)
4	8032 (21.69)	4894 (21.60)	12,926 (21.66)
5 (highest)	12,251 (33.09)	7347 (32.43)	19,598 (32.84)
Unknown	530 (1.43)	210 (0.93)	740 (1.24)

Membership and Platform Attrition

The percentage of members with paid membership and active platform usage declined gradually over the first 24 weeks (Figure 2). At week 12, 94.56% (56,438/59,686) of members had a paid membership, and at week 13, this decreased to 69.65% (41,573/59,686) of members. By week 14, 65.34% (39,000/59,686) of the original membership base had a paid membership, which continued to decrease to 41.48% (24,756/59,686) of members at 24 weeks. In terms of platform usage, 83.80% (46,893/55,958) of members were using the platform (that is any logged data recorded) after 3 weeks, 52.03% (29,115/55,958) of remaining members were using the platform at week 12, and 26.59% (14,880/55,958) of those remaining were still using the platform by week 24 (Figure 2).

Figure 2. Program membership and platform usage attrition over 24 weeks for members of the CSIRO Total Wellbeing Diet Online (n=59,686).



Weight Loss

The primary analysis using last observation carried forward indicated an average weight loss for all members of 2.8 kg or 3.1% of starting body weight. Stayers lost an average of 4.9 kg (5.3% of starting body weight) compared to starters, who lost an average of 1.6 kg (1.7% of starting body weight). The starting BMI of members classified as starters and stayers was similar to that at the start of the program. Stayers who were classified as overweight when they joined the program lost 4.9% of their starting body weight, compared to 5.5% for those classified as class 1 obese, 5.6% as class 2, and 5.9% as class 3. Stayers living in the most disadvantaged areas lost 5.8% of their starting body weight compared to 4.9% for those in areas of least disadvantage (Table 2).

Overall, 24.37% (14,546/59,686) of all members achieved a weight loss of 5% or more of their starting body weight. Almost half (11,082/22,658, 48.91%) of the stayers lost 5% or more of their starting body weight compared to 9.36% (3464/37,028) of starters (Table 3). Among stayers, 33.43% (7575/22,658) achieved a weight loss of 5% to <10%, and 15.48% (3507/22,658) achieved a weight loss of 10% or more of their starting body weight. The proportion of members losing >10% of their starting weight was highest for males (lost 10% or more of their starting body weight: 1012/4503, 22.47%), members classified as class 3 obese at baseline (445/2263, 19.66%), and those living in the most disadvantaged areas (409/2322, 17.61%; Table 3).

Table 4 shows the weight status of members at the time they joined and at the end of the program. Overall, 14.27% (2701/18,930) of all members who were classified as overweight when they joined the program had achieved normal weight by the end. Among stayers, this proportion was 24.59% (1911/7770). Across all levels of obesity, 24% (3180/13,319) of stayers who were classified as obese when they joined the program were no longer classified as obese at the end. More specifically, among stayers who were classified as class 1 obese at baseline, 41.39% (3065/7405) had shifted to a lower weight status category at the end of the program (ie, overweight or normal weight). In addition, 47.86% (1748/3652) and 36.78% (832/2262) of stayers who were classified as class 2 and 3, respectively, achieved a lower weight status at the end of the program.

Table 2. Weight loss in kilograms and percentage of body weighta in starters (n=37,028), stayers (n=22,658), and all members (n=59,686) on the CSIRO Total Wellbeing Diet Online.

Member characteristics	Starters, mean (SD)		Stayers, mean (SD)		All members, mean (SD)	
	Weight loss (kg)	% body weight	Weight loss (kg)	% body weight	Weight loss (kg)	% body weight
Total	1.60 (3.51)	1.73 (3.36)	4.87 (5.60)	5.25 (5.56)	2.84 (4.69)	3.07 (4.65)
Sex						
Male	2.28 (4.44)	2.14 (3.79)	6.85 (6.26)	6.51 (5.45)	4.20 (5.75)	3.98 (5.05)
Female	1.47 (3.27)	1.65 (3.26)	4.38 (5.31)	4.94 (5.54)	2.55 (4.38)	2.87 (4.54)
Age (years)						
18-30	1.64 (2.70)	1.80 (2.99)	4.48 (6.22)	4.94 (6.29)	2.50 (4.30)	2.75 (4.50)
31-50	1.61 (3.24)	1.72 (3.18)	4.46 (5.70)	4.75 (5.63)	2.58 (4.46)	2.76 (4.42)
51-70	1.61 (3.71)	1.75 (3.41)	5.12 (5.56)	5.54 (5.51)	3.06 (4.88)	3.31 (4.78)
>71	1.43 (4.55)	1.54 (4.59)	5.39 (4.61)	5.99 (4.76)	3.20 (4.98)	3.53 (5.16)
Starting BMI category						
Normal weight	0.76 (2.41)	1.15 (3.90)	2.38 (3.03)	3.58 (4.51)	1.22 (2.70)	1.83 (4.23)
Overweight	1.40 (2.71)	1.78 (3.51)	3.87 (4.05)	4.91 (5.11)	2.42 (3.54)	3.07 (4.51)
Obese	1.92 (4.00)	1.88 (3.19)	5.72 (6.31)	5.66 (5.84)	3.41 (5. 36)	3.34 (4.78)
Class 1	1.75 (2.54)	1.90 (2.27)	5.11 (5.31)	5.54 (5.63)	3.11 (4.24)	3.37 (4.52)
Class 2	1.93 (3.14)	1.84 (3.01)	5.90 (6.16)	5.64 (5.80)	3.44 (4.93)	3.29 (4.67)
Class 3	2.41 (7.14)	1.85 (4.42)	7.45 (8.77)	5.94 (6.55)	4.24 (8.14)	3.34 (5.65)
Missing/invalid	0.56 (3.15)	0.56 (2.40)	7.61 (14.70)	7.53 (8.81)	0.82 (4.37)	0.82 (3.17)
Socioeconomic status						
1 (lowest)	1.78 (4.77)	1.82 (4.13)	5.58 (6.10)	5.80 (5.76)	3.26 (5.64)	3.36 (5.20)
2	1.71 (2.68)	1.83 (2.58)	5.20 (5.77)	5.53 (5.56)	3.04 (4.47)	3.24 (4.37)
3	1.70 (4.14)	1.81 (3.99)	5.00 (5.46)	5.34 (5.58)	2.98 (4.97)	3.19 (4.98)
4	1.59 (3.71)	1.72 (3.58)	4.97 (5.56)	5.37 (5.50)	2.87 (4.79)	3.10 (4.75)
5 (highest)	1.49 (2.85)	1.66 (2.89)	4.39 (5.27)	4.87 (5.40)	2.57 (4.18)	2.86 (4.31)
Unknown	0.83 (1.66)	0.92 (1.83)	3.49 (9.02)	3.81 (8.15)	1.59 (5.14)	1.74 (4.78)

[a] Average weight loss in kilograms and percentage of starting body weight were calculated using last observation carried forward.

Table 3. Percentage of the sample within weight loss categoriesa,b for starters (n=37,028), stayers (n=22,658), and all members (N=59,686) on the CSIRO Total Wellbeing Diet Online.

Member characteristics	Starters, n (%)				Stayers, n (%)				All members, n (%)			
	Weight gain	0%-<5% lost	5%-<10% lost	≥10% lost	Weight gain	0%-<5% lost	5%-<10% lost	≥10% lost	Weight gain	0%-<5% lost	5%-<10% lost	≥10% lost
Total	2227 (6.01)	31337 (84.63)	3148 (8.5)	316 (0.85)	2518 (11.11)	9058 (39.98)	7575 (33.43)	3507 (15.48)	4745 (7.95)	40395 (67.68)	10723 (17.97)	3823 (6.41)
Sex												
Male	284 (4.58)	5006 (80.69)	815 (13.14)	99 (1.6)	317 (7.04)	1489 (33.07)	1685 (37.42)	1012 (22.47)	601 (5.61)	6495 (60.66)	2500 (23.35)	1111 (10.38)
Female	1943 (6.3)	26331 (85.42)	2333 (7.57)	217 (0.7)	2201 (12.12)	7569 (41.69)	5890 (32.44)	2495 (13.74)	4144 (8.46)	33900 (69.21)	8223 (16.79)	2712 (5.54)
Age (years)												
18-30	160 (6.92)	1896 (82.01)	219 (9.47)	37 (1.6)	132 (13.15)	377 (37.55)	349 (34.76)	146 (14.54)	292 (8.81)	2273 (68.55)	568 (17.13)	183 (5.52)
31-50	1055 (6.69)	13287 (84.28)	1294 (8.21)	130 (0.82)	1136 (13.85)	3359 (40.95)	2584 (31.5)	1124 (13.7)	2191 (9.14)	16646 (69.45)	3878 (16.18)	1254 (5.23)
51-70	919 (5.37)	14558 (85.09)	1497 (8.75)	134 (0.78)	1158 (9.68)	4741 (39.64)	4081 (34.12)	1980 (16.56)	2077 (7.15)	19299 (66.39)	5578 (19.19)	2114 (7.27)
71 +	93 (5.05)	1596 (86.64)	138 (7.49)	15 (0.81)	92 (6.17)	581 (38.97)	561 (37.63)	257 (17.24)	185 (5.55)	2177 (65.32)	699 (20.97)	272 (8.16)
Starting BMI category												
Normal weight	252 (6.68)	3280 (86.98)	227 (6.02)	12 (0.32)	226 (15.34)	732 (49.69)	427 (28.99)	88 (5.97)	478 (9.12)	4012 (76.51)	654 (12.47)	100 (1.91)
Overweight	710 (6.36)	9374 (83.91)	987 (8.84)	100 (0.9)	895 (11.49)	3162 (40.61)	2682 (34.45)	1047 (13.45)	1605 (8.47)	12536 (66.13)	3669 (19.35)	1147 (6.05)
Obese	1238 (5.94)	17529 (84.04)	1896 (9.09)	196 (0.94)	1391 (10.43)	5140 (38.54)	4444 (33.32)	2361 (17.70)	2629 (7.69)	22669 (66.29)	6340 (18.54)	2557 (7.48)
Class 1	676 (6.18)	9087 (83.08)	1073 (9.81)	101 (0.92)	804 (10.84)	2797 (37.72)	2511 (33.86)	1304 (17.58)	1480 (8.06)	11884 (64.75)	3584 (19.53)	1405 (7.66)
Class 2	359 (6.03)	5012 (84.15)	534 (8.97)	51 (0.86)	374 (10.23)	1434 (39.21)	1237 (33.83)	612 (16.74)	733 (7.63)	6446 (67.06)	1771 (18.42)	663 (6.90)
Class 3	203 (5.12)	3430 (86.49)	289 (7.29)	44 (1.11)	213 (9.41)	909 (40.17)	696 (30.76)	445 (19.66)	416 (6.68)	4339 (69.66)	985 (15.81)	489 (7.85)
Missing /invalid	24 (1.99)	1136 (94.27)	37 (3.07)	8 (0.66)	2 (4.35)	17 (36.96)	17 (36.96)	10 (21.74)	26 (2.08)	1153 (92.17)	54 (4.32)	18 (1.44)
Socioeconomic status												
1 (lowest)	222 (6.06)	3055 (83.36)	351 (9.58)	37 (1.01)	226 (9.73)	874 (37.64)	813 (35.01)	409 (17.61)	448 (7.48)	3929 (65.63)	1164 (19.44)	446 (7.45)
2	312 (5.59)	4728 (84.75)	496 (8.89)	43 (0.77)	346 (10.06)	1310 (38.09)	1214 (35.3)	569 (16.55)	658 (7.3)	6038 (66.95)	1710 (18.96)	612 (6.79)
3	405 (5.81)	5876 (84.29)	619 (8.88)	71 (1.02)	474 (10.66)	1729 (38.89)	1520 (34.19)	723 (16.26)	879 (7.7)	7605 (66.61)	2139 (18.74)	794 (6.95)
4	502 (6.25)	6821 (84.92)	647 (8.06)	62 (0.77)	537 (10.97)	1904 (38.9)	1673 (34.18)	780 (15.94)	1039 (8.04)	8725 (67.5)	2320 (17.95)	842 (6.51)
5 (highest)	762 (6.22)	10376 (84.7)	1011 (8.25)	102 (0.83)	899 (12.24)	3146 (42.82)	2300 (31.31)	1002 (13.64)	1661 (8.48)	13522 (69)	3311 (16.89)	1104 (5.63)
Unknown	24 (4.53)	481 (90.75)	24 (4.53)	1 (0.19)	36 (17.1)	95 (45.2)	55 (26.2)	24 (11.4)	60 (8.11)	576 (77.84)	79 (10.68)	25 (3.38)

[a] Average weight loss in kilograms and percentage of starting body weight were calculated using last observation carried forward.

[b] Proportion of the sample within each category of weight loss: weight gain, 0 to less than 5% of starting body weight lost, 5% to less than 10% of

starting body weight lost, and greater than or equal to 10% of starting body weight lost.

Table 4. Shift in body weight status category as a percentage of starting weight status category for starters (n=37,028), stayers (n=22,658), and all members (N=59,686) in the CSIRO Total Wellbeing Diet Online.

Starting weight status category	Final weight status category, n (%)				
	Normal weight	Overweight	Class 1 obese	Class 2 obese	Class 3 obese
Starters					
Normal weight (n=3767)	3716 (98.65)	45 (1.19)	4 (0.11)	1 (0.03)	1 (0.03)
Overweight (n=11,160)	790 (7.08)	10,297 (92.27)	68 (0.61)	1 (0.01)	4 (0.04)
Class 1 obese (n=10,933)	9 (0.08)	1572 (14.38)	9286 (84.94)	65 (0.59)	1 (0.01)
Class 2 obese (n=5955)	4 (0.07)	5 (0.08)	1053 (17.68)	4871 (81.80)	22 (0.37)
Class 3 obese (n=3966)	5 (0.13)	6 (0.15)	7 (0.18)	502 (12.66)	3446 (86.89)
Stayers					
Normal weight (n=1468)	1391 (94.75)	76 (5.18)	1 (0.07)	0 (0.00)	0 (0.00)
Overweight (n=7770)	1911 (24.59)	5688 (73.20)	166 (2.14)	2 (0.03)	3 (0.04)
Class 1 obese (n=7405)	56 (0.76)	3009 (40.63)	4223 (57.03)	113 (1.53)	4 (0.05)
Class 2 obese (n=3652)	11 (0.30)	92 (2.52)	1645(45.04)	1851 (50.71)	52 (1.42)
Class 3 obese (n=2262)	4 (0.18)	8 (0.35)	64 (2.83)	756 (33.42)	1430 (63.22)
All members					
Normal weight (n=5235)	5107 (97.55)	121 (2.31)	5 (0.10)	1 0.02)	1 (0.02)
Overweight (n=18,930)	2701 (14.27)	15985 (84.44)	234 (1.24)	3 (0.02)	7 (0.04)
Class 1 obese (n=18,338)	65 (0.35)	4581 (24.98)	13,509 (73.67)	178 (0.97)	5 (0.03)
Class 2 obese (n=9607)	15 (0.16)	97 (1.01)	2698 (28.08)	6723 (69.98)	74 (0.77)
Class 3 obese (n=6228)	9 (0.14)	14 (0.22)	71 (1.14)	1258 (20.20)	4876 (78.29)

Platform Usage

Overall members used the platform on 29.8% of their membership days, and this was higher in stayers than starters (46.5% vs 19.6%, respectively). The most commonly used platform features were the weigh-in (used by all members), food diary (52,828/59,686, 88.51% of members), and menu plans (51,718/59,686, 86.65% of members). The total number of days (irrespective of membership length) members were active on the features was highest for the food diary (15.10 days), menu plans (13.17 days), and weigh-in feature (12.27 days), but activity per day of membership was highest for the food diary and menu plan (used approximately 2 out of every 3 days; Table 5).

The usage per day of membership was higher in stayers for all features. On average, stayers used the menu plans and food diary once per day, while starters used these less than once every 2 days. Stayers used the weigh-in once per week, while starters used this feature less than once every fortnight (Table 5).

Table 5. Platform feature usage for starters (n=37,028), stayers (n=22,658) and all members (n=59,686) of the CSIRO Total Wellbeing Diet Online.

Platform feature	Usage, n (%)[a]	Total days active, mean (SD)[b]	Active days, mean (SD)[c]	Activity per day, mean (SD)[d]
Starters (n= 37,028)				
Overall	37,028 (100.00)	22.02 (24.34)	19.61 (28.70)	0.98 (2.09)
Weigh-in	36,965 (99.83)	4.63 (4.98)	4.30 (8.50)	0.04 (0.08)
Food diary	30,900 (83.45)	6.76 (9.27)	6.05 (10.30)	0.37 (0.89)
Menu plan	30,637 (82.74)	6.73 (10.40)	5.70 (9.92)	0.48 (1.17)
Exercise plan	19,654 (53.08)	1.35 (2.58)	1.25 (2.87)	0.03 (0.07)
Program content	17,203 (46.46)	1.01 (1.93)	0.92 (2.18)	0.01 (0.03)
Forum	13,319 (35.97)	0.89 (2.46)	0.83 (2.94)	0.02 (0.07)
Food search	9538 (25.76)	0.63 (1.99)	0.57 (2.69)	0.03 (0.44)
Stayers (n=22,658)				
Overall	22,658 (100.00)	90.89 (113.54)	46.50 (46.59)	2.30 (3.12)
Weigh-in	22,615 (99.81)	24.75 (35.18)	12.79 (18.44)	0.13 (0.18)
Food diary	21,926 (96.77)	28.73 (45.73)	14.97 (18.21)	0.95 (1.58)
Menu plan	21,079 (93.03)	23.69 (39.93)	11.58 (15.16)	1.03 (1.71)
Exercise plan	15,992 (70.58)	3.77 (8.40)	2.14 (4.84)	0.05 (0.13)
Program content	16,021 (70.71)	2.89 (5.69)	1.62 (3.00)	0.02 (0.06)
Forum	12,679 (55.96)	4.14 (19.89)	1.99 (6.25)	0.04 (0.24)
Food search	11,372 (50.19)	2.92 (9.21)	1.41 (3.66)	0.08 (0.31)
All members (n=59,686)				
Overall	59,686 (100.00)	48.17 (79.86)	29.82 (38.80)	1.48 (2.61)
Weigh-in	59,579 (99.82)	12.27 (24.09)	7.52 (13.82)	0.08 (0.14)
Food diary	52,828 (88.51)	15.10 (31.00)	9.43 (14.51)	0.59 (1.23)
Menu plan	51,718 (86.65)	13.17 (27.21)	7.93 (12.50)	0.69 (1.43)
Exercise plan	35,650 (59.73)	2.27 (5.69)	1.59 (3.77)	0.03 (0.1)
Program content	33,227 (55.67)	1.73 (3.93)	1.18 (2.54)	0.02 (0.05)
Forum	25,999 (43.56)	2.13 (12.50)	1.27 (4.53)	0.03 (0.16)
Food search	20,914 (35.04)	1.50 (5.99)	0.89 (3.12)	0.05 (0.40)

[a] Percentage of the sample who used the platform or feature at any time during their membership.

[b] Total days active refers to the number of unique days a member used the platform irrespective of whether usage occurred multiple times per day.

[c] Percent active days was calculated as total days active divided by number of days of membership expressed as a percentage.

[d] Activity per day was calculated as the total user activity divided by the number of days of membership.

Weight Loss Across Levels of Platform Usage

Table 6 shows the average weight loss by levels of platform usage. High usage was characterized as viewing the menu plans 2-3 times per day, making 2 food diary entries per day, searching for foods once per week, recording a weight about once per week, and viewing the exercise plans and the forum a little less than once per week. Weight loss in kilograms and percentage body weight increased with increasing levels of platform usage. Overall, members with the highest usage lost 5.3% of their starting body weight compared to 0.9% for those with the lowest usage. Stayers who used the platform most lost 6.6% of their starting body weight compared to 3.0% for those with the lowest usage. Among stayers, 1 in 5 in the highest category of usage lost 10% or more of their starting body weight and 2 in 5 lost 5%-<10% of their starting body weight.

Table 6. Weight loss by level of platform usage for starters (n=37,028), stayers (n=22,658), and all members (N=59,686) in the CSIRO Total Wellbeing Diet Online.

Levels of platform usage	Weight loss, mean (SD)[a]		Weight loss in categories, n (%)[b]			
	Weight, kg	% body weight	Weight gain	0%-<5% lost	5%-<10% lost	≥10% lost
Starters						
Quintile 1: lowest usage (n=10,471)	0.55 (2.78)	0.57 (3.12)	582 (5.56)	9610 (91.78)	237 (2.26)	42 (0.40)
Quintile 2 (n=8337)	1.40 (4.20)	1.45 (3.74)	627 (7.52)	7151 (85.77)	501 (6.01)	58 (0.70)
Quintile 3 (n=7290)	1.86 (3.75)	2.01 (3.28)	496 (6.80)	6047 (82.95)	677 (9.29)	70 (0.96)
Quintile 4 (n=6398)	2.31 (3.04)	2.54 (2.84)	357 (5.58)	5162 (80.68)	816 (12.75)	63 (0.98)
Quintile 5: highest usage (n=4532)	2.99 (3.05)	3.33 (2.91)	165 (3.64)	3367 (74.29)	917 (20.23)	83 (1.83)
Stayers						
Quintile 1: lowest usage (n=1466)	2.88 (6.15)	2.98 (6.30)	360 (24.56)	670 (45.70)	299 (20.40)	137 (9.35)
Quintile 2 (n=3600)	4.03 (5.86)	4.26 (5.70)	572 (15.89)	1583 (43.97)	987 (27.42)	458 (12.72)
Quintile 3 (n=4649)	4.45 (5.49)	4.72 (5.45)	606 (13.04)	2012 (43.28)	1436 (30.89)	595 (12.80)
Quintile 4 (n=5539)	4.82 (5.30)	5.21 (5.37)	591 (10.67)	2284 (41.23)	1853 (33.45)	811 (14.64)
Quintile 5: highest usage (n=7404)	5.98 (5.41)	6.56 (5.22)	389 (5.25)	2509 (33.89)	3000 (40.52)	1506 (20.34)
All members						
Quintile 1: lowest usage (n=11,937)	0.84 (3.46)	0.87 (3.75)	942 (7.89)	10280 (86.12)	536 (4.49)	179 (1.50)
Quintile 2 (n=11,937)	2.19 (4.91)	2.30 (4.61)	1199 (10.04)	8734 (73.17)	1488 (12.47)	516 (4.32)
Quintile 3 (n=11,939)	2.87 (4.68)	3.07 (4.46)	1102 (9.23)	8059 (67.50)	2113 (17.70)	665 (5.57)
Quintile 4 (n=11,937)	3.47 (4.42)	3.78 (4.41)	948 (7.94)	7446 (62.38)	2669 (22.36)	874 (7.32)
Quintile 5: highest usage (n=11,936)	4.84 (4.84)	5.33 (4.75)	554 (4.64)	5876 (49.23)	3917 (32.82)	1589 (13.31)

[a] Average weight loss in kilograms and percentage of starting body weight were calculated using last observation carried forward.

[b] Proportion of the sample within each category of weight loss: weight gain, 0 to less than 5% of starting body weight lost, 5% to less than 10% of starting body weight lost, and greater than or equal to 10% of starting body weight lost.

Predictors of Weight Loss

On the basis of multiple linear regression, member characteristics (sex, age, socioeconomic status, starting BMI) and program characteristics (duration of membership and usage of the platform per day) accounted for 29.4% of the variance associated with the percentage total weight loss ($F_{6,58434}$=920.01; *P*<.001). All predictors were significant, with platform usage being the strongest predictor of weight loss (β=.263; *P*<.001) and higher usage per day being associated with greater weight loss. Sex (β=–.086; *P*<.001) and starting BMI (β=.062; *P*<.001) were the next 2 strongest predictors of percentage body weight lost, with men and those with a higher starting BMI losing more weight.

Sensitivity Analysis

Starters, by definition, did not have a weight entered into the platform at 12 weeks or more; therefore, we used baseline observation to carry forward their assumed weight loss to zero. The average weight loss for stayers was 4.9 kg. Using the baseline-observation-carried-forward approach, we assumed the average weight loss for all members was 1.8 kg. Multiple linear regression showed similar results to those of the primary analysis. The model accounted for 31.5% of the variance associated with percentage total weight loss, and platform usage was the strongest predictor of weight loss (β=.272; *P*<.001).

Discussion

Principal Results

This study examined the data generated through the CSIRO Total Wellbeing Diet Online, a commercial weight loss program, to determine its reach and effectiveness over its first 5 years of use. This analysis included usage data from almost all (59,686/61,164, 97.58%) of the individuals who had signed up between October 2014 to September 2019. The results indicated that the average weight loss across all members was 3.1% of their starting body weight but was considerably higher (5.3% of starting body weight) in the 22,658 members who completed the full 12-week program or more (37.96% of all members analyzed). This magnitude of weight loss was consistent with similar programs [17,19] and greater than that reported for some other web-based commercial programs [10]. However, it was less than that of commercial programs that use a more intensive, one-to-one consultation model [28].

Our finding that almost 1 in 2 members who completed the CSIRO Total Wellbeing Diet Online program lost 5% or more of their starting body weight has considerable public health relevance, as a loss of 5% of body weight is considered clinically significant in relation to reductions in risk of comorbidities [29]. This result is also similar to the findings of a 6-month German-based commercial weight loss program, although the evaluation of this program used a much smaller sample (N=479) [30]. Furthermore, overweight and obese status are related to the development of several chronic conditions, which place a huge burden on the health care system [1]. This study was also able to demonstrate that about 40% of members with obesity who stayed on the program were able to shift down a weight status category, with about 24% no longer classified as obese when they finished the program. Moreover, approximately 25% who were classified as overweight at the start of the program achieved a normal weight at the completion of the program. The increased risk of morbidity and mortality from many chronic conditions means programs that can shift people out of the higher-risk, obese weight status category can confer significant potential gains in health and quality of life for these individuals [31]. If the results demonstrated here can be achieved at an even greater scale, the potential social and economic benefits will be extremely significant.

One of the most important findings from this study was that members who were living in the most disadvantaged areas were one of the most successful groups on the program based on percentage of total weight loss. About 10% of members were living in areas considered to be among the most disadvantaged in Australia, and their average weight loss was 5.8% of starting body weight, and about 53% lost 5% percent or more of their starting body weight. Although this may be partly attributable to the higher starting BMI in this group, it is nevertheless an encouraging finding given that individuals of lower socioeconomic status are at increased risk of weight gain and obesity [32,33] but tend to have lower rates of uptake and completion of weight loss programs [34]. There are many unique challenges that have been cited for disadvantaged groups, including literacy, insufficient time, and lack of motivation [32]. It has also been suggested that individuals with lower income are less likely to recognize their unhealthy weight and therefore less likely to attempt weight loss than are higher income individuals [35]. It is likely that motivations for joining a weight loss program may vary across groups of differing socioeconomic status [32]. Overall, the CSIRO Total Wellbeing Diet Online program appears to be an effective weight loss program for individuals in lower socioeconomic groups, representing a simple, evidence-based, and effective intervention to address the public health challenge of being overweight or obese.

Platform Usage and Weight Loss

Participant adherence is known to be a primary determinant of program success [36], and this study, like others [11,37], showed that higher engagement with an online platform and more frequent usage of program features was associated with greater early weight loss and greater total weight loss. Members with the highest platform usage lost twice as much weight than did those with the lowest levels of usage. A review of technology-driven weight loss interventions suggests that self-monitoring, feedback and communication, social support, use of a structured program, and use of an individually tailored program are factors that could facilitate weight loss [38]. Opportunity for self-monitoring through entries in a diary [30,37,39,40] and engagement with social forums have been shown to be associated with greater weight loss and weight loss maintenance in web-based interventions [11,37,41]. On this program, members with the highest platform usage were using the food diary to record dietary intake about twice per day. They were also monitoring their body weight more closely by entering a weight value into the platform once per week. Consistent self-monitoring of weight is crucial to successful weight loss and weight loss maintenance [42] and is central to behavioral weight loss programs [43].

The specific features of a program are important and can influence engagement and weight loss. Different technologies may lend themselves to different features. The focus of this study was on a web-based program, but commercial weight loss programs delivered through mobile phone apps are also common. One of the most popular publicly available apps for weight loss is Noom Coach, with tens of millions of installs worldwide. An evaluation of Noom Coach used data from a very select group who had recorded their data at least twice per month for 6 consecutive months (n=36,000). In this highly motivated subsample, who were engaged and used the app in this way, 23% of users achieved a weight loss of 10% or more over the 6-month period examined [44]. For the CSIRO Total Wellbeing Diet Online, 20% (4532/22,658) of an engaged subsample (that is the highest platform usage users who stayed on the program for at least 3 months) achieved a weight loss of 10% or more. However, it would be interesting to see a larger-scale study with a greater proportion of total Noom users to better understand the weight loss a typical user can expect when embarking on the program. Mobile phone apps can offer different features, such as push notifications and real time information, which may provide additional support that is not available in a web-based program. An increasing number of technologies are being used in parallel, and programs are available across multiple platforms to offer users greater flexibility and choice. As the complexity of program delivery increases, it is vital that robust evaluations are conducted to determine the elements of programs, the components of their delivery, and the combinations of these that are associated with successful weight loss.

Achieving sufficient engagement and user retention is a challenge. For commercial weight loss programs, high attrition rates of up to 70% are common [28,45], and nonusage attrition is also high on web-based weight loss programs [46]. It is important to look at both program attrition and attrition in terms of technology usage because it is possible that people stop using the platform features before they formally drop out [46]. Platform usage could be thought of as the equivalent of treatment dose in medical studies, and so, while the optimal dose for web-based interventions is unclear, strategies to improve engagement need to be determined. Engagement and retention on a program might be related to user characteristics such as starting weight, as people who are heavier have more weight to lose to reach their weight loss goal and therefore are willing to stay on a program for longer [28]. In this study, starters and stayers had a similar starting BMI, but stayers might have been more intrinsically motivated to adhere to the program and achieve weight loss. A greater understanding of user behaviors that predict engagement and whether this varies for different user groups is needed to develop strategies to increase engagement and drive more successful weight loss outcomes. For example, posting a profile picture has been shown to predict higher engagement with an online weight loss program [47], and more frequent input of evening meal information is the most important behavior for weight loss in a commercial weight loss app [48]. If more specific behaviors like these were recognized, then they could be incorporated and specifically promoted to help achieve higher levels of engagement and weight loss.

Rapid initial weight loss has been shown to predict greater longer-term weight loss [30,49], and while this study showed higher usage of the platform was associated with greater early weight loss, further exploration could determine specific behavioral patterns that support early success. Data were not available for members once they had dropped out (and thus our results represent the best-case scenario), and the reasons for drop-out and nonplatform usage are also worthy of further investigation. It is unknown how weight loss is maintained once membership has ceased, but the achievement of long-term weight loss or weight loss maintenance following completion of the program is the true goal.

Strengths and Limitations

A key strength of the current study is that a broad set of inclusion criteria was applied for analysis, and data from all members were used in reporting the weight loss results, which is not always the case for commercial programs [9,10,44]. Weight loss was conducted using self-reported values with the last weight entered into the platform being carried forward. As we split the sample into starters and stayers and presented the data separately, the weight losses reported for stayers are likely to be reflective of an “average” member that joins and finishes the program and not a selected subsample. Although this is a retrospective analysis, it represents a very large cohort of people who had chosen to join a commercial weight loss program and followed the program in their own everyday context. However, a limitation associated with this design is that the study included only the member characteristics that were collected as part of the registration process. Although a range of personal demographic data was included, other factors, such as past dieting history or motivation, were not available. Platform usage provided an indication of engagement, while viewing the menu plans and use of the food diary may relate to compliance with the eating plan. However, more objective and validated measures of dietary intake would better represent dietary program compliance.

Conclusions

To our knowledge, this 5-year evaluation of the CSIRO Total Wellbeing Diet Online represents one of the most comprehensive evaluations of a commercially available, online weight loss program and has provided insights into predictors of weight loss success among members. Health care practitioners and individuals require evidenced-based weight management programs that have been shown to achieve meaningful weight loss. The findings suggest that this program is effective for weight loss, particularly for members who finish the program and are active in using the online platform and tools provided. Importantly, members with a higher starting BMI and those from more disadvantaged socioeconomic areas were among those who lost the most weight.

Acknowledgments

We acknowledge the team at Digital Wellness for their support in maintaining the CSIRO Total Wellbeing Diet Online platform, retrieving the data, and assisting in the data management prior to analysis.

Conflicts of Interest

This study was cofunded by CSIRO and Digital Wellness. Digital Wellness manages the CSIRO Total Wellbeing Diet Online under a commercial agreement. Authors from CSIRO conducted the analysis and were responsible for the interpretation of results and preparation of the manuscript. Digital Wellness was not involved in the analysis or preparation of results. Digital Wellness has approved the manuscript for publication.

Multimedia Appendix 1

Mean (SD) starting body weight and body mass index of starters (n=37,028), stayers (n=22,658) and all members (n=59,686) at the time of joining the CSIRO Total Wellbeing Diet Online.

[DOCX File , 16 KB - jmir_v23i6e20981_app1.docx]

References

1. A picture of overweight and obesity in Australia. Australian Institute of Health and Welfare. Canberra; 2017. URL: https://www.aihw.gov.au/getmedia/172fba28-785e-4a08-ab37-2da3bbae40b8/aihw-phe-216.pdf.aspx?inline=true [accessed 2021-02-04]

2. National Health Survey: first results. Australian Bureau of Statistics. Canberra, Australia; 2017. URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/4364.0.55.001 [accessed 2021-02-03]
3. Healthy diet. World Health Organization. URL: https://www.who.int/topics/diet/en/ [accessed 2021-10-28]
4. Stubbs RJ, Morris L, Pallister C, Horgan G, Lavin JH. Weight outcomes audit in 1.3 million adults during their first 3 months' attendance in a commercial weight management programme. BMC Public Health 2015 Sep 10;15:882 [FREE Full text] [doi: 10.1186/s12889-015-2225-0] [Medline: 26359180]
5. Weight loss services in Australia: Australia Industry (ANZSIC) Report S9512. IBISWorld. URL: https://my.ibisworld.com/download/au/en/industry/1704/1/0/pdf [accessed 2021-01-20]
6. Atallah R, Filion KB, Wakil SM, Genest J, Joseph L, Poirier P, et al. Long-term effects of 4 popular diets on weight loss and cardiovascular risk factors: a systematic review of randomized controlled trials. Circulation: Cardiovascular Quality and Outcomes 2014 Nov 11;7(6):815-827. [doi: 10.1161/circoutcomes.113.000723]
7. Johnston BC, Kanters S, Bandayrel K, Wu P, Naji F, Siemieniuk RA, et al. Comparison of weight loss among named diet programs in overweight and obese adults: a meta-analysis. JAMA 2014 Sep 03;312(9):923-933. [doi: 10.1001/jama.2014.10397] [Medline: 25182101]
8. Gudzune KA, Doshi RS, Mehta AK, Chaudhry ZW, Jacobs DK, Vakil RM, et al. Efficacy of commercial weight-loss programs. Ann Intern Med 2015 Apr 07;162(7):501. [doi: 10.7326/m14-2238]
9. Lowe MR, Kral TVE, Miller-Kovach K. Weight-loss maintenance 1, 2 and 5 years after successful completion of a weight-loss programme. Br J Nutr 2007 Nov 28;99(4):925-930. [doi: 10.1017/s0007114507862416]
10. 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]
11. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
12. Ryan K, Dockray S, Linehan C. A systematic review of tailored eHealth interventions for weight loss. Digit Health 2019;5:2055207619826685 [FREE Full text] [doi: 10.1177/2055207619826685] [Medline: 30783535]
13. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J Med Internet Res 2017 Jun 26;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]
14. Arem H, Irwin M. A review of web-based weight loss interventions in adults. Obes Rev 2011 May;12(5):e236-e243 [FREE Full text] [doi: 10.1111/j.1467-789X.2010.00787.x] [Medline: 20804523]
15. Manzoni GM, Pagnini F, Corti S, Molinari E, Castelnuovo G. Internet-based behavioral interventions for obesity: an updated systematic review. Clin Pract Epidemiol Ment Health 2011 Mar;7:19-28 [FREE Full text] [doi: 10.2174/1745017901107010019] [Medline: 21552423]
16. Tsai AG, Wadden TA. Systematic review: an evaluation of major commercial weight loss programs in the United States. Ann Intern Med 2005 Jan 04;142(1):56-66. [doi: 10.7326/0003-4819-142-1-200501040-00012] [Medline: 15630109]
17. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. A 12-week commercial web-based weight-loss program for overweight and obese adults: randomized controlled trial comparing basic versus enhanced features. J Med Internet Res 2012;14(2):e57 [FREE Full text] [doi: 10.2196/jmir.1980] [Medline: 22555246]
18. Thomas JG, Raynor HA, Bond DS, Luke AK, Cardoso CC, Foster GD, et al. Weight loss in weight watchers online with and without an activity tracking device compared to control: a randomized trial. Obesity (Silver Spring) 2017 Jun;25(6):1014-1021. [doi: 10.1002/oby.21846] [Medline: 28437597]
19. Womble LG, Wadden TA, McGuckin BG, Sargent SL, Rothman RA, Krauthamer-Ewing ES. A randomized controlled trial of a commercial internet weight loss program. Obes Res 2004 Jun;12(6):1011-1018 [FREE Full text] [doi: 10.1038/oby.2004.124] [Medline: 15229342]
20. Clifton P, Abbey M, Noakes M, Nestel P. Body fat distribution is a determinant of the lipoprotein response to dietary fat and cholesterol. Atherosclerosis 1994 Sep;109(1-2):148. [doi: 10.1016/0021-9150(94)93604-8]
21. Noakes M, Clifton P. Dietary fats in energy restricted diet effects on cardiovascular risk factors. Circulation 1999;100(18):e.
22. Farnsworth E, Luscombe N, Noakes M, Wittert G, Argyiou E, Clifton P. Effect of a high-protein, energy-restricted diet on body composition, glycemic control, and lipid concentrations in overweight and obese hyperinsulinemic men and women. Am J Clin Nutr 2003 Jul;78(1):31-39. [doi: 10.1093/ajcn/78.1.31] [Medline: 12816768]
23. Luscombe ND, Clifton PM, Noakes M, Farnsworth E, Wittert G. Effect of a high-protein, energy-restricted diet on weight loss and energy expenditure after weight stabilization in hyperinsulinemic subjects. Int J Obes Relat Metab Disord 2003 May;27(5):582-590. [doi: 10.1038/sj.ijo.0802270] [Medline: 12704402]
24. Noakes M, Clifton P. The CSIRO Total Wellbeing Diet. Camberwell, Victoria: Penguin Group (Australia); 2006:2.
25. Wyld B, Harrison A, Noakes M. The CSIRO Total Wellbeing Diet Book 1: sociodemographic differences and impact on weight loss and well-being in Australia. Public Health Nutr 2010 Dec;13(12):2105-2110. [doi: 10.1017/S136898001000073X] [Medline: 20392311]

26. Total Wellbeing Diet. Commonwealth Scientific and Industrial Research Organisation. URL: https://www.totalwellbeingdiet.com [accessed 2021-05-01]
27. Australian Bureau of Statistics. Socio-Economic Indexes for Areas. URL: https://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa [accessed 2017-03-15]
28. Finley CE, Barlow CE, Greenway FL, Rock CL, Rolls BJ, Blair SN. Retention rates and weight loss in a commercial weight loss program. Int J Obes (Lond) 2007 Feb;31(2):292-298. [doi: 10.1038/sj.ijo.0803395] [Medline: 16755283]
29. Magkos F, Fraterrigo G, Yoshino J, Luecking C, Kirbach K, Kelly S, et al. Effects of moderate and subsequent progressive weight loss on metabolic function and adipose tissue biology in humans with obesity. Cell Metab 2016 Apr 12;23(4):591-601 [FREE Full text] [doi: 10.1016/j.cmet.2016.02.005] [Medline: 26916363]
30. Postrach E, Aspalter R, Elbelt U, Koller M, Longin R, Schulzke J, et al. Determinants of successful weight loss after using a commercial web-based weight reduction program for six months: cohort study. J Med Internet Res 2013;15(10):e219 [FREE Full text] [doi: 10.2196/jmir.2648] [Medline: 24126250]
31. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
32. Harvey JR, Ogden DE. Obesity treatment in disadvantaged population groups: where do we stand and what can we do? Prev Med 2014 Nov;68:71-75 [FREE Full text] [doi: 10.1016/j.ypmed.2014.05.015] [Medline: 24878585]
33. Ball K, Crawford D. Socioeconomic status and weight change in adults: a review. Soc Sci Med 2005 May;60(9):1987-2010. [doi: 10.1016/j.socscimed.2004.08.056] [Medline: 15743649]
34. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]
35. Johnston DW, Lordan G. Weight perceptions, weight control and income: an analysis using British data. Econ Hum Biol 2014 Jan;12:132-139 [FREE Full text] [doi: 10.1016/j.ehb.2013.02.004] [Medline: 23578515]
36. Adherence to long-term therapies: evidence for action. World Health Organization. 2003. URL: https://www.who.int/chp/knowledge/publications/adherence_report/en/ [accessed 2021-02-03]
37. 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;14(6):e173 [FREE Full text] [doi: 10.2196/jmir.2156] [Medline: 23234759]
38. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938 [FREE Full text] [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]
39. 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;15(1):e11 [FREE Full text] [doi: 10.2196/jmir.2195] [Medline: 23322819]
40. Neve M, Morgan PJ, Collins CE. Weight change in a commercial web-based weight loss program and its association with website use: cohort study. J Med Internet Res 2011;13(4):e83 [FREE Full text] [doi: 10.2196/jmir.1756] [Medline: 21993231]
41. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Couto SAP, Meira W, et al. Factors associated with weight change in online weight management communities: a case study in the LoseIt Reddit community. J Med Internet Res 2017 Jan 16;19(1):e17 [FREE Full text] [doi: 10.2196/jmir.5816] [Medline: 28093378]
42. Butryn ML, Phelan S, Hill JO, Wing RR. Consistent self-monitoring of weight: a key component of successful weight loss maintenance. Obesity (Silver Spring) 2007 Dec;15(12):3091-3096 [FREE Full text] [doi: 10.1038/oby.2007.368] [Medline: 18198319]
43. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. Journal of the American Dietetic Association 2011 Jan;111(1):92-102. [doi: 10.1016/j.jada.2010.10.008]
44. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6:34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
45. Volkmar FR, Stunkard AJ, Woolston J, Bailey RA. High attrition rates in commercial weight reduction programs. Arch Intern Med 1981 Mar;141(4):426-428. [Medline: 7212881]
46. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. J Med Internet Res 2010;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]
47. Freyne J, Saunders I, Brindal E, Berkovsky S, Smith G. Factors associated with persistent participation in an online diet intervention. 2012 Presented at: CHI Conference on Human Factors in Computing Systems; May 10, 2012; Austin, TX, United States. [doi: 10.1145/2212776.2223805]
48. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo J, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6:34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]

49. Hemmingsson E, Johansson K, Eriksson J, Sundström J, Neovius M, Marcus C. Weight loss and dropout during a commercial weight-loss program including a very-low-calorie diet, a low-calorie diet, or restricted normal food: observational cohort study. The American Journal of Clinical Nutrition 2012;96(5):953-961. [doi: 10.3945/ajcn.112.038265]

Abbreviations

ANOVA: analysis of variance
CSIRO: Commonwealth Scientific and Industrial Research Organisation
RCT: randomized controlled trial
SEIFA: Socio-Economic Indexes for Areas

Edited by R Kukafka; submitted 05.07.20; peer-reviewed by R Krukowski; comments to author 27.07.20; revised version received 04.11.20; accepted 19.04.21; published 07.06.21.

Please cite as:
Hendrie GA, Baird DL, Brindal E, Williams G, Brand-Miller J, Muhlhausler B
Weight Loss and Usage of an Online Commercial Weight Loss Program (the CSIRO Total Wellbeing Diet Online) Delivered in an Everyday Context: Five-Year Evaluation in a Community Cohort
J Med Internet Res 2021;23(6):e20981
URL: https://www.jmir.org/2021/6/e20981
doi:10.2196/20981
PMID:34096869

Original Paper

Frequency of Self-Weighing and Weight Change: Cohort Study With 10,000 Smart Scale Users

Anna-Leena Vuorinen[1,2*], PhD; Elina Helander[3*], PhD; Julia Pietilä[3], PhD; Ilkka Korhonen[3], PhD

[1] VTT Technical Research Centre of Finland, Tampere, Finland
[2] Health Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland
[3] Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
[*] these authors contributed equally

Corresponding Author:
Anna-Leena Vuorinen, PhD
VTT Technical Research Centre of Finland
P.O. Box 1300
Tampere, FI-33101
Finland
Phone: 358 408485966
Email: anna-leena.vuorinen@vtt.fi

Abstract

Background: Frequent self-weighing is associated with successful weight loss and weight maintenance during and after weight loss interventions. Less is known about self-weighing behaviors and associated weight change in free-living settings.

Objective: This study aimed to investigate the association between the frequency of self-weighing and changes in body weight in a large international cohort of smart scale users.

Methods: This was an observational cohort study with 10,000 randomly selected smart scale users who had used the scale for at least 1 year. Longitudinal weight measurement data were analyzed. The association between the frequency of self-weighing and weight change over the follow-up was investigated among normal weight, overweight, and obese users using Pearson's correlation coefficient and linear models. The association between the frequency of self-weighing and temporal weight change was analyzed using linear mixed effects models.

Results: The eligible sample consisted of 9768 participants (6515/9768, 66.7% men; mean age 41.5 years; mean BMI 26.8 kg/m2). Of the participants, 4003 (4003/9768, 41.0%), 3748 (3748/9768, 38.4%), and 2017 (2017/9768, 20.6%) were normal weight, overweight, and obese, respectively. During the mean follow-up time of 1085 days, the mean weight change was –0.59 kg, and the mean percentage of days with a self-weigh was 39.98%, which equals 2.8 self-weighs per week. The percentage of self-weighing days correlated inversely with weight change, r=–0.111 (P<.001). Among normal weight, overweight, and obese individuals, the correlations were r=–0.100 (P<.001), r=–0.125 (P<.001), and r=–0.148 (P<.001), respectively. Of all participants, 72.5% (7085/9768) had at least one period of ≥30 days without weight measurements. During the break, weight increased, and weight gains were more pronounced among overweight and obese individuals: 0.58 kg in the normal weight group, 0.93 kg in the overweight group, and 1.37 kg in the obese group (P<.001).

Conclusions: Frequent self-weighing was associated with favorable weight loss outcomes also in an uncontrolled, free-living setting, regardless of specific weight loss interventions. The beneficial associations of regular self-weighing were more pronounced for overweight or obese individuals.

(J Med Internet Res 2021;23(6):e25529) doi:10.2196/25529

KEYWORDS

self-monitoring; self-weighing; weight change, weight loss, normal weight, overweight, obese, temporal weight change

Introduction

Self-monitoring of body weight is the cornerstone of behavioral weight loss interventions [1]. Increased awareness gained through self-weighing is expected to trigger a self-evaluation process that involves interpretation of weight data against one's goal and subsequently results in weight loss–promoting actions [2,3]. Regular self-weighing is consistently associated with weight loss and weight loss maintenance as well as weight gain prevention [2-7]. There is also evidence of a dose-response relationship with more frequent self-weighing and greater weight loss [8-11]. Currently, less is known about self-weighing practices and their effectiveness independent of the context of weight loss interventions.

To date, most studies have investigated self-weighing behaviors through self-reported methods. Self-weighing frequency is often retrospectively reported and evaluated in predefined categories (eg, “daily weighing,” “weekly weighing,” or “monthly weighing”) [9,12,13]. Some studies have determined self-weighing frequency from actual self-weighing data, but further reduced it to a single aggregated frequency value [10,14,15]. Although valuable in research, self-reported data collection methods have their limitations, including selective reporting and susceptibility to recall bias [16,17]. Moreover, single aggregated categories may not reflect long-term weighing practices, as self-weighing frequency tends to change or decline over time [18-20].

Self-monitoring of body weight has become easier and increasingly popular with modern smart scales that automatically transmit weight data to network servers that users can access online or through smart phone applications. As a result, date-stamped and time-stamped weight record data accumulate from a large number of users, providing unique insights into behavioral patterns of weight-tracking individuals. Advantages of using these data in health research include objectivity and accuracy of the information they provide [16,21]. In addition, they provide longitudinal weight measurement data, which are rarely, although increasingly, obtained or used in research. However, while being a strength, self-weighing with smart scales is done by individuals in their daily lives (ie, in uncontrolled, free-living settings), which introduces challenges for the analytics and interpretation of the data. These include the possibility of weight measurement outliers, daily variation, and inadequate background information, among others [21-25].

In our previous work, we demonstrated a dose-response relationship between self-weighing frequency and weight change and an increased risk for weight gain when weight monitoring was stopped [8]. However, the generalization of the findings was limited due to the small sample size (n=40) and the sample consisting of overweight individuals engaged in a workplace health promotion program. The objective of this study was to investigate self-weighing patterns and their association with weight change in a large international sample of smart scale users in their free-living settings. We divided the sample into 3 groups to further study the differences among normal weight, overweight, and obese individuals.

Methods

Design and Data Source

The data used in the current study were collected from 10,000 anonymous Withings (Withings Ltd, Paris, France) smart scale users randomly retrieved from the company’s smart scale user database between May 2009 and June 2015. A user was defined as a person who had access to a smart scale and had a user account in which his or her measurements were linked. In the connection of setting up the user account, the user consented to his or her data being used for research purposes by accepting the Terms and Conditions of Withings [26]. For each measurement, the scale automatically saved the weight record and the date stamp and time stamp of the measurement, and the information was further sent to the network servers. Multiple users could share the same scale; however, they were advised to create an individual user account. The scale automatically detected to which user the weight measurement belonged. Users were able to remove weight observations or attach unknown observations to their account manually through a web-based application interface. The data used in the study were pseudonymized before data retrieval and contained all weight measurements attached to a single user account.

In order to study both long-term and short-term weight changes, the smart scale users who fulfilled the following inclusion criteria by June 2015 were eligible for random sampling: engaged with self-weighing for at least 1 year (based on a time difference between the first and the last weight measurements) and had ≥30 weight measurements linked to a user account.

Background information of smart scale users included the following self-reported measures: age in years, height, gender, and location information (ie, country and time zone) that were automatically recorded.

Study Subjects

In addition to the inclusion criteria used for data extraction, smart scale users who fulfilled the following criteria were included in the current study: age between 18 and 100 years at the time of the first weight measurement, height between 1.40 m and 2.10 m (defined to exclude possible errors in height records that would lead to erroneous BMI values), BMI ≥18.5 kg/m^2 based on the weight measurements made during the first follow-up week.

Data Processing

As the weight measurement data were not collected through a structured protocol, but instead accumulated in a free-living setting, the data required preprocessing to exclude possible outlying measurements and multiple intraday measurements from the current analyses. Outliers may arise from situations such as different users using the same devices or due to external, uncontrolled influences such as exceptionally heavy clothing or carrying an object during self-weighing. We identified and removed the outliers using the following steps. First, all measurements below 30 kg were removed. Second, the outlying measurements were identified with the median absolute deviation method (window=30 samples, threshold=4) and subsequently removed. Third, weight measurements resulting in an unrealistic weight change from the preceding measurement were identified and excluded using a specific algorithm: A weight change was considered unrealistic if it was voted 3 times as an outlier by the 2 preceding and 2 successive measurements, with each measurement encompassing 1 vote. A vote was defined as a difference between the 2 consecutive weight measurements that exceeded the threshold of 3% daily plus 0.1% times day difference between the measurements. The measurements were removed iteratively by removing 1 observation at a time, starting from the outlier that had the highest number of votes and then the highest total sum of absolute weight difference. The process was repeated until no outliers were detected. After removing the outliers, the minimum weight values for each day were identified, and other day duplicates were excluded from the analyses to reduce diurnal variation of weight. Finally, if more than 10 original daily minimum weight values were removed as a result of the outlier detection algorithm (compared to before any processing) or if less than 30 observations were left after preprocessing, the user was excluded from the final analyses.

Variables

Follow-Up Time

The individual follow-up time for each user was defined as the time difference between the first and the last weight measurements.

Weight Change

Weight change during the follow-up period was defined as a difference between the mean weight of the last 7 days and the mean weight of the first 7 days (*initial weight*), divided by the initial weight. A similar approach was used by Zheng et al [19]. Negative and positive weight differences indicate weight loss and weight gain, respectively.

Temporal Weight Change

Temporal weight change per day was calculated by dividing weight change between 2 consecutive measurements by the corresponding day difference between the measurements.

BMI

Height was retrieved from the Withings research database. The baseline BMI was calculated from the self-reported height and the initial weight. Individuals were categorized into 3 categories based on BMI thresholds: normal weight (18.5 kg/m^2 to <25.0 kg/m^2), overweight (25.0 kg/m^2 to <30.0 kg/m^2), and obese (≥30.0 kg/m^2).

Frequency of Self-Weighing

The self-weighing frequency was defined as a percentage of follow-up days with a weight measurement recorded. We also calculated the number of weight measurements per week defined by dividing the number of all weight measurements with the number of follow-up weeks to be used as a proxy for weekly self-weighing intensity.

For the analysis of temporal weight changes, we used a categorical self-weighing frequency that was defined based on the number of days between the 2 consecutive measurements. The categories were based on commonly used self-weighing categories. The categorization was also motivated by the fact that the majority of measurements were done on a weekly basis or more frequently, and thus, the distribution was strongly skewed. The self-weighing categories were as follows: daily — 0 days between 2 consecutive weight measurements; every other day — 1 day between 2 consecutive weight measurements; 2-4 times per week — 2 days between 2 consecutive weight measurements; 1-2 times per week — 3-7 days between 2 consecutive weight measurements; every other week — 8-14 days between 2 consecutive weight measurements; monthly — 14-29 days between 2 consecutive weight measurements; less than monthly — ≥30 days between consecutive measurements.

Break in Self-Weighing

A break in self-weighing was defined as ≥30 days between 2 consecutive weight measurements. The weight change during the break was calculated as a difference between the last preceding weight measurement and the first weight measurement after the break. If the user had multiple breaks, the mean weight change and the mean duration of all breaks were calculated.

Time of the Day and the Weekday

To control time effect, weight measurements were each categorized based on their time stamp to either a morning (measurement made between 5 am and 1 pm) or an evening (measurement made between 1 pm and the 5 am the next day) measurement. We further defined a combination variable accounting for time stamps of the present and the previous weight measurement with 4 possible categories: morning-morning, morning-evening, evening-evening, evening-morning. This variable was used in the analyses of temporal weight changes. Similarly, as weekday is shown to affect weight variation [24,27], we defined the weekday of the weight measurement and used the variable in the analyses of temporal weight change.

Statistical Analyses

Participant characteristics are presented using means (SDs) or frequencies and percentages. Characteristics were compared across the 3 BMI groups using a Chi-square test or analysis of variance or Kruskall Wallis test if the distributions were strongly skewed. Similar comparisons across the BMI groups were made among participants who had a break in self-weighing.

The Pearson correlation coefficient (r) between weight change (kg) and the frequency of self-weighing, measured as a percentage of follow-up days with a weight measurement, was calculated for the whole study population and separately for each BMI group. We then built linear models with an interaction term between the BMI group and the frequency of self-weighing to investigate whether the association was different across the BMI groups. The adjusted models were further performed including the following covariates: sex, age, initial BMI, and the duration of the follow-up.

In the secondary analyses, we focused on temporal weight changes and analyzed the association between temporal weight change and the corresponding day difference by using linear mixed effect (LME) models. Day difference between 2 consecutive weight measurements was added in the model as a fixed effect and participants as a random effect. Each participant was allowed to have a subject-specific intercept and a slope. The adjusted analyses included the following covariates: sex, age, initial BMI, weekday, and time of the day combination. Temporal weight change was also modeled using the categorized self-weighing frequency variable as a dependent variable and further adding the interaction between the self-weighing category and BMI group in the model as a fixed effect.

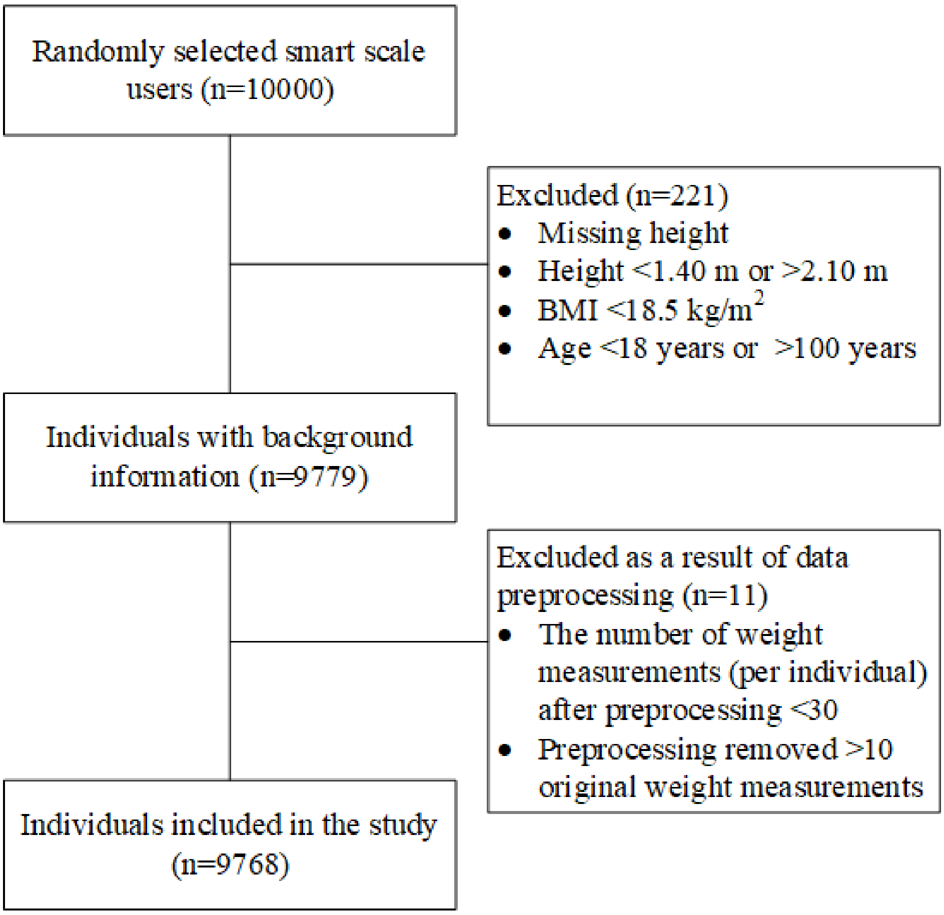
Statistical analyses were carried out with R version 3.6.2 [28] using dplyr [29] and nlme [30] packages. An alpha level of 0.05 was used in all statistical analyses, and all statistical tests were two-sided.

Results

Participant Inclusion and Background Characteristics

The participant inclusion and data preprocessing flow is depicted in Figure 1. Of the 10,000 smart scale users, 9768 (97.7%) were included in the study. In total, participants recorded 4,230,928 weight measurements.

Figure 1. Participant inclusion and data prepreprocessing.



Participants' background characteristics and self-weighing–related information are summarized in Table 1. The majority (6515/9768, 66.7%) were male. The mean age was 41.52 years, and the mean BMI was 26.78 kg/m^2. Of all participants, 4003 (4003/9768, 42.0%) were of normal weight, 3748 (3748/9768, 38.4%) were overweight, and 2017 (2017/9768, 20.7%) were obese at the beginning of the follow-up. The proportion of male participants was higher in the overweight and obese groups than among the normal weight group (2969/3748, 79.2%; 1551/2017, 76.9%; 1995/4003, 49.8%, respectively; *P*<.001). Of the included participants, 4837 (4837/9768, 49.5%) were from Europe, 3744 (3744/9768, 38.3%) from the Americas, 887 (887/9758, 9.1%) from Asia, and the rest (300/9768, 3.0%) from other continents. The most represented countries were the United States (3332/9768, 34.1%), Germany (1230/9768, 12.6%), France (811/9768, 8.3%), Japan (600/9768, 6.1%), and the United Kingdom (585/9768, 6.0%).

The mean duration of the follow-up was 1085 days, which did not differ across the 3 BMI groups (*P*=.35). The median number of self-weighs was 352, which equals 2.80 weight measurements per week. The highest self-weighing intensity was seen among overweight individuals; however, the differences between the groups were small, although statistically significant, varying from 38.99% among obese individuals to 40.86% among overweight individuals (*P*=.006). The mean percentage days with a weight measurement was 39.98%.

Table 1. Background characteristics and self-weighing–related information for smart scale users.

Characteristic	All participants (n=9768)	Normal weight group (n=4003)	Overweight group (n=3748)	Obese group (n=2017)	Statistic for the difference between the BMI groups	*P* value
Gender (male), n (%)	6515 (66.7)	1995 (49.8)	2969 (79.2)	1551 (76.9)	χ^2_2=871.15	<.001
Age (years), mean (SD)	41.52 (11.08)	40.07 (10.86)	42.33 (11.07)	42.88 (11.21)	$F_{2,9765}$=60.24	<.001
Weight (kg), mean (SD)	82.01 (19.03)	67.01 (9.96)	84.74 (9.86)	106.71 (17.48)	N/A[a]	N/A
BMI (kg/m^2), mean (SD)	26.78 (5.01)	22.55 (1.68)	27.20 (1.42)	34.38 (4.32)	N/A	N/A
Follow-up time (days), mean (SD)	1085 (420.17)	1086 (422.65)	1090 (419.16)	1073 (417.08)	$F_{2,9765}$=1.05	.35
Total number of weight measurements, median (IQR)	352.00 (197.00 to 591.00)	345.00 (190.00 to 588.50)	371.00 (213.75 to 604.00)	338.00 (189.00 to 571.00)	KW_2[b]=18.59	<.001
Percentage of days with a weight measurement, mean (SD)	39.98 (22.80)	39.65 (23.41)	40.86 (22.24)	38.99 (22.54)	$F_{2,9765}$=5.09	.006
Weight measurements per week, mean (SD)	2.80 (1.60)	2.78 (1.64)	2.86 (1.56)	2.73 (1.58)	$F_{2,9765}$=5.09	.006
Weight change (kg), mean (SD)	–0.59 (6.99)	0.78 (4.38)	–0.58 (5.96)	–3.35 (11.01)	$F_{2,9765}$=246.00	<.001
Weight change (%), mean (SD)	–0.37 (7.64)	1.20 (6.57)	–0.68 (7.09)	–2.93 (9.61)	$F_{2,9765}$=209.4	<.001
Maximum weight change (kg), median (IQR)	–4.97 (–8.52 to –2.76)	–3.49 (–5.42 to –2.03)	–5.81 (–8.97 to –3.38)	–9.33 (–15.73 to –4.99)	KW_2=1827.00	<.001

[a] N/A: not available because no statistical test was performed to test differences between the groups.

[b] KW: Kruskal-Wallis.

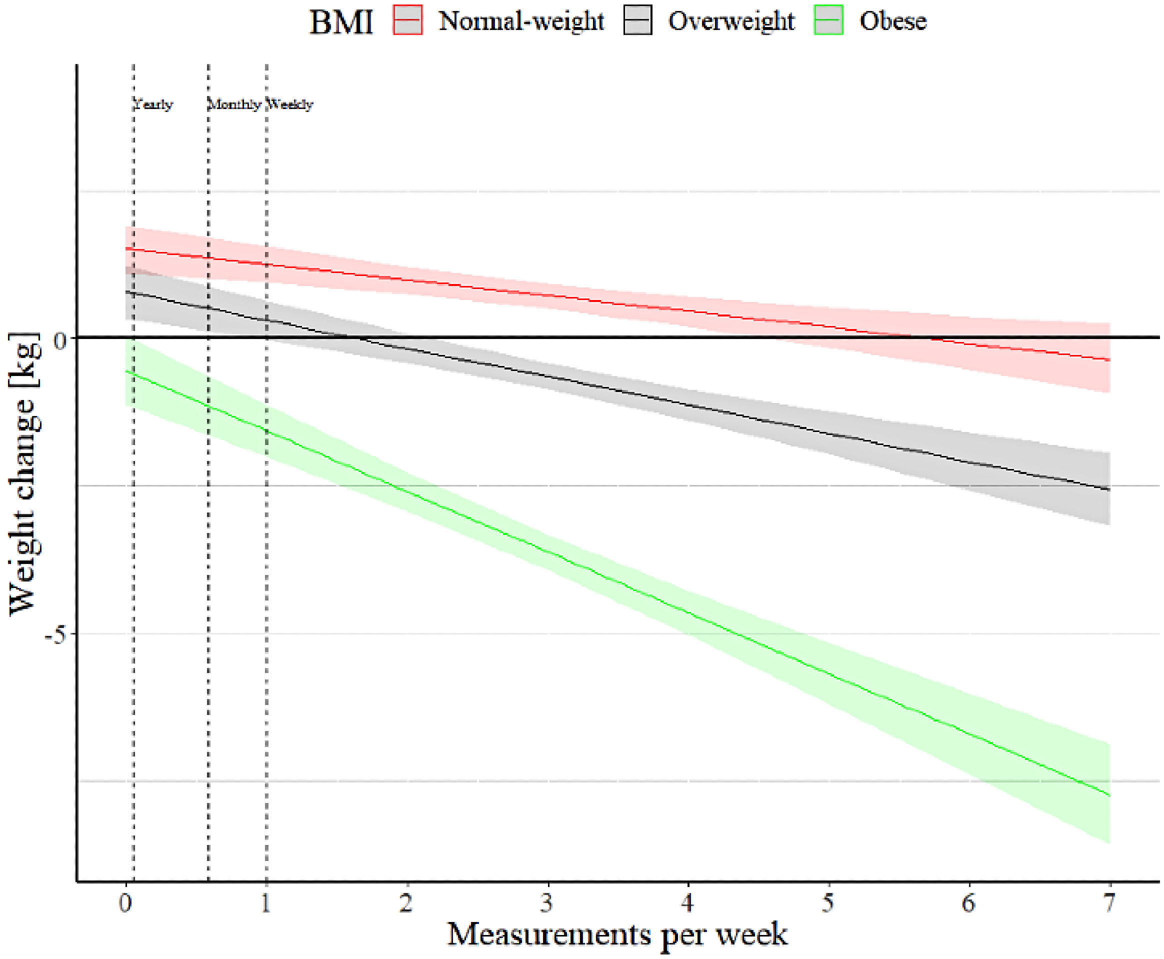
Frequency of Self-Weighing and Weight Change

During the mean follow-up of 1085 days, the participants lost 0.59 kg of their initial weight (Table 1). The mean percentage weight change was –0.37%. Weight changes varied between the BMI groups, with an increase of 0.78 kg in the normal weight group and decreases of –0.58 kg and –3.35 kg in the overweight and obese groups, respectively.

There was a low, statistically significant, correlation between the frequency of self-weighing, measured as the percentage of self-weighing days, and weight change (r=–0.111, P<.001). The correlation was stronger in the highest BMI groups: r=–0.100 (P<.001) for normal weight, r=–0.125 (P<.001) for overweight, and r=–0.148 (P<.001) for obese. The linear model, including the interaction term between the BMI group and the frequency of self-weighing, showed a statistically significant interaction for the overweight and obese groups and the frequency of self-weighing (β=–0.015, P=.03 and β=–0.053, P<.001, respectively); see Table S1 in Multimedia Appendix 1 for all model coefficients. The result indicates that, when compared to the normal weight group, the association between the frequency of self-weighing and weight change was different among obese and overweight individuals. Similar parameter estimates were obtained from the adjusted models (β=–0.017, P=.01 for the overweight group and β=–0.057, P<.001 for the obese group).

Figure 2 shows the weight change (kg) in the 3 BMI groups predicted by the weekly weighing frequency. A slightly steeper slope was seen in the obese group as compared to the normal weight group illustrating the interaction. A higher frequency of self-weighing was associated with bigger weight loss in the obese group. The magnitude of the interaction was smaller in the overweight group. The figure shows that weighing oneself at least once a week was associated with negative weight change in the obese group, whereas in the overweight and normal weight groups, the confidence intervals overlap the 0 line or remain above it. It seems that a self-weighing frequency of 2-3 times per week, or more frequently, was associated with negative weight change in the overweight group. However, in the normal weight group, the upper limit of the 95% confidence interval was not below 0 at all.

Figure 2. Weight change during the study follow-up predicted by the number of weight measurements per week.

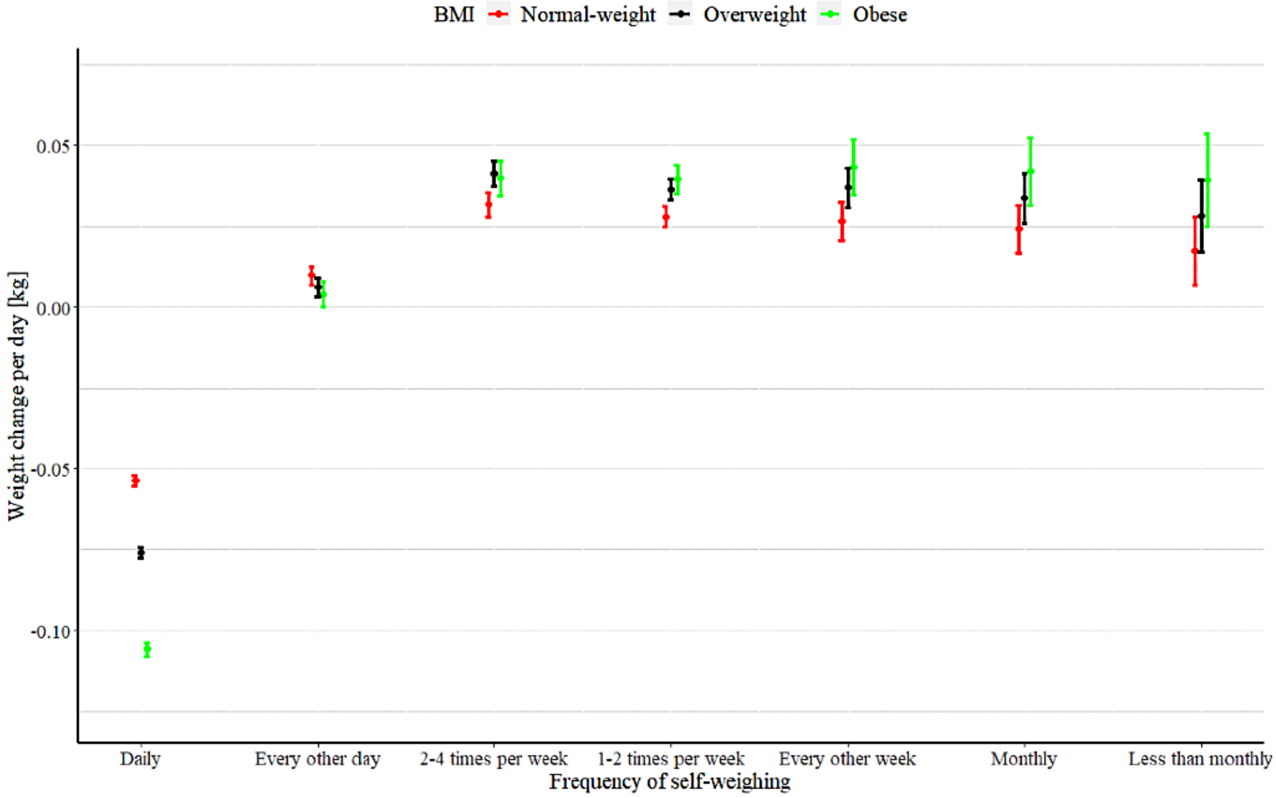


Temporal Weight Change and Self-Weighing Frequency

After calculating the temporal weight change variable, we excluded measurements that exceeded the threshold of 20% weight change per day (n=35). The number of individuals in the analyses remained at 9768. The LME model showed that the number of days between 2 consecutive weight measurements was a significant predictor of temporal weight change (β=0.001, *P*<.001); weight increased by the number of days between the weight measurements. The adjusted model did not alter the parameter estimate (β=0.001, *P*<.001).

The mean weight changes per day in the self-weighing frequency categories were as follows: daily, –0.058 kg; every other day, 0.012 kg; 2-4 times per week, 0.036 kg; 1-2 times per week, 0.028 kg; every other week, 0.026 kg; monthly, 0.022 kg; less than monthly, 0.016 kg. Temporal weight changes in each category were significantly different from the "daily" value, which was used as a reference category (Multimedia Appendix 1, Table S2). Figure 3 shows the interactions between self-weighing categories and the BMI groups. The parameter estimates of the model are shown in Multimedia Appendix 1 Table S3. Only daily self-weighing was associated with a negative temporal weight change per day in all BMI groups.

Figure 3. Weight change per day predicted by the interaction between the frequency of self-weighing and BMI group.



Breaks in Self-Weighing

A majority of the participants (7085/9768, 72.5%) had at least one break in self-weighing of ≥30 days during the follow-up (Table 2). In the normal weight, overweight, and obese groups, 73.5% (2941/4003), 70.7% (2648/3748), and 74.2% (1496/2017), respectively, had at least one break. The median number of breaks was 3 with an average duration of 58.83 days.

During the break, the participants gained 0.85 kg on average. The median increase was bigger in the obese group (1.37 kg) than in the overweight (0.93 kg) and normal weight (0.58 kg) groups (*P*<.001). The Spearman correlation coefficients between the duration of the break and weight change (kg) were ρ=0.153 (*P*<.001) among all participants and ρ=0.090 (*P*<.001), ρ=0.161 (*P*<.001), and ρ=0.245 (*P*<.001) in the normal weight, overweight, and obese groups, respectively.

Table 2. The numbers of breaks in self-weighing, their durations, and the subsequent weight change among all study participants who had a break from self-weighing and in the 3 BMI groups.

Characteristic	All who had a break from self-weighing (n=7085)	Normal weight (n=2941)	Overweight (n=2648)	Obese (n=1496)	Kruskal Wallis statistic for the difference between the BMI groups	*P* value
Number of breaks, median (IQR)	3 (2 to 5)	3 (2 to 5)	3 (1 to 5)	3 (2 to 6)	8.48 (df=2)	.01
Duration of the break (days), median (IQR)	58.83 (44.00 to 86.30)	58.33 (44.00 to 85.00)	57.75 (43.62 to 85.55)	61.20 (45.36 to 90.00)	11.44 (df=2)	.003
Weight change (kg), median (IQR)	0.85 (0.03 to 1.92)	0.58 (–0.13 to 1.46)	0.93 (0.06 to 1.99)	1.37 (0.38 to 2.61)	247.32 (df=2)	<.001
Weight change (%), median (IQR)	1.05 (0.04 to 2.30)	0.87 (–0.18 to 2.20)	1.10 (0.08 to 2.32)	1.30 (0.36 to 2.46)	58.54 (df=2)	<.001

Discussion

Principal Findings

This study extends the well-established relationship between frequent self-weighing and favorable weight loss outcomes beyond specific weight loss interventions by using a large international sample of 9768 long-term smart scale users, tracking their weight in uncontrolled, free-living settings. We found that the intensity of self-weighing was inversely associated with weight change, and the beneficial associations of frequent self-weighing were more pronounced in obese and overweight individuals as compared to normal weight individuals. Daily self-weighing in particular was associated with weight loss.

During the mean follow-up time of 1085 days, the participants measured their weight 2.8 times per week, which is notably higher than what is reported in the general population of the United States [15]. However, a comparable self-weighing intensity was reported in the Heart eHealth (HeH) study [19], a prospective observational cohort study on health behaviors and cardiovascular risk. Similarly, the HeH study reflects free-living behaviors, as study participants did not receive any recommendations for self-weighing practices, but those who owned smart scales connected them to the HeH study account.

While the self-weighing frequency was rather high, more than two-thirds of the study participants also had at least one longer break in self-weighing (≥30 days) during the follow-up. The break in self-weighing was associated with increased weight, with a dose-response relationship with the length of the break: The longer the break, the more weight tended to increase. Moreover, weight gain following the break in self-weighing was largest among obese individuals (+1.37 kg) and overweight individuals (+0.93 kg) when compared to normal weight individuals (+0.58 kg), while the average length of the break was rather similar between all 3 BMI groups (varying between 58 days and 61 days). Similar findings were also reported by VanWormer et al [12], who found that monthly self-weighers actually gained weight during the 24-month follow-up. These results suggest that regular self-weighing might be of particular importance in terms of promoting weight loss for individuals with excess weight, possibly by helping them attain weight management–promoting behaviors.

The correlation between frequent self-weighing and weight loss supports the beneficial effect of weight monitoring and, as such, is in line with other studies [3,8,9,11,13]. The association seemed to be stronger in individuals who needed it most, namely obese and overweight individuals, as seen earlier [12]. The motivation for weight tracking in individuals who do not have excess weight is likely different, as they may not pursue weight loss, which might at least partly explain the lower correlation in the normal weight group. It should be noted that the correlation coefficients were generally very low, which indicates that self-weighing explains relatively little of the weight change variation. Other factors, such as long-term goals, motivation for self-weighing, and behavioral factors as well as diet and physical activity, may have stronger effects on weight loss efforts, although weight-loss–promoting practices are important. It has also been suggested that the consistency of self-weighing is more important than the frequency of self-weighing [31].

As self-weighing practices tend to vary and decline over time [13,20,32], we analyzed temporal weight changes using consecutive weight measurements and the corresponding day differences to investigate the short-term associations. These results were consistent with aggregated data showing a significant relationship between weight change and the corresponding day difference. The higher the number of days between self-weighs, the more weight increased. In these analyses, we used weight change per day as an independent variable, which adjusted the weight change by the number of days between the measurements to control for the possibility of bigger weight changes during longer breaks. When using the self-weighing categories, we found that only daily self-weighing was associated with weight loss in all BMI groups, whereas weighing oneself every other day or less frequently was associated with unchanged or increased weight. Interestingly, weighing every other day was associated with increased weight in normal weight and overweight groups, whereas in obese individuals, it was associated with unchanged weight. Consistent with our findings, it has been shown that daily weighing is associated with greater weight loss [33], even when compared to weighing oneself 5 days of the week [11]. However, not all studies have reported the superiority of daily self-weighing over weekly monitoring [5]. The positive impact of daily weighing vs weighing oneself every other day or on a weekly basis in our observational study should be interpreted with caution. It is possible that the differences between daily and every-other-day categories reflect an ongoing weight loss process rather than daily self-weighing being remarkably more effective than weighing oneself every other day.

Limitations

The findings of this study should be interpreted in the light of the following limitations. First, as an observational study, a causal relationship between self-weighing and weight change cannot be established. Efforts at self-weighing could promote weight loss or be a consequence of favorable weight change. In fact, a recent study by Frie et al [25] found that declining motivation for weight control and possible difficulties with making use of negative weight feedback, the Ostrich effect, precede a break in self-weighing. Thus, it might not be the break in self-weighing that increases the risk of weight gain, but that the decline in self-weighing behavior is a result of weight gain. In line with this, Tanenbaum et al [34] showed that calorie intake on a given day increased the risk of not self-weighing the following day; in another study, a reduction in BMI was independently associated with greater engagement with self-weighing [21]. In light of those studies, the findings of the current study may also indicate that the relationship describes the process of weight loss rather than promotes it.

Second, participants' background information was limited to demographic variables, preventing us from controlling for important variables, such as goals for weight management or the underlying motivation for using the smart scale. For example, among normal weight individuals, it might be that they did not pursue weight loss but rather weight (loss) maintenance, which might explain a weaker correlation seen in this subgroup. On the other hand, obese individuals might be more likely to use the scale for weight loss purposes.

The third limitation is the selection bias. Our sample consisted of individuals who were predominantly men (66.7%), had excess weight (BMI of 26.8 kg/m^2), and, importantly, were engaged with regular self-monitoring of weight for a longer time period. The sample is not representative of the general population; Sperrin et al [21], who used comparable data from the same company, showed that the traits of those self-weighing individuals differed from the general British population. Participants' characteristics in our study were quite similar to the study by Sperring et al [21], although they required only one weight measurement per subject and used data from a single country. Another study found that individuals using smart scales in wireless settings were mostly well educated, white, and male [19]. Moreover, due to the inclusion criteria we applied for sampling, our cohort may not be representative of smart scale users, as those who used the scale for a shorter period (less than a year) and made fewer than 30 measurements were excluded. Nevertheless, this was a large international cohort including almost 10,000 individuals from 109 different countries who provided self-weighed measurements for approximately 3 years. These individuals were not participating in a specific weight loss program, but the data reflect self-weighing behaviors in real-world settings. We consider the overrepresentation of men as a strength of this study. Typically, participants in weight management interventions are women. Thus, the findings of this study shed light on self-weighing patterns and weight loss attempts by men, who are less investigated. Moreover, together with other smart scale–based studies, the findings indicate that including technology to support weight loss may have the potential to appeal to men to participate in weight control trials.

Fourth, measuring longitudinal changes in self-weighing behaviors remain a challenge. Aggregating data over multiple days and weeks, as we did in the first analysis, does not account for changing behavioral patterns. In an attempt to tackle the challenge, we investigated short-term weight changes by analyzing consecutive weight measurements. However, these results may have been affected by the underlying trend of the weight loss process; individuals who lose weight may be more likely to weigh themselves daily, and thus, even 1-day breaks from the daily self-weighing may appear as short-term weight gains. As there were some differences in results obtained with these 2 analysis approaches, future studies should note that the length of the follow-up might affect the results when investigating how the frequency of self-weighing affects weight loss. Another alternative for the analysis of similar data to ours would be to identify slots with unchanged self-weighing pattern as was done by Frie et al [25] and to analyze associated weight changes.

Conclusions

This study showed that frequent self-weighing is associated with favorable weight loss outcomes also in an uncontrolled, free-living setting, regardless of specific weight loss interventions. The positive association of regular self-weighing was more pronounced for individuals who needed it most, namely overweight and obese individuals. We found that daily self-weighing in particular was associated with weight loss, whereas breaks of 30 days or longer were associated with increased weight. However, although, there was a dose-response relationship between self-weighing intensity and weight change, the correlation remained low, indicating that self-weighing explains only a small fraction of the variations in weight. Yet, our results underscore the evidence that missing self-weighing data do not occur randomly and might be a sign of risk for weight gain.

Acknowledgments

We would like to thank Angela Chieh from Withings for her contribution to the study design and data collection. We are also grateful to Withings for the collaboration and kindly providing data for this research. The work was funded by VTT Technical Research Centre of Finland.

Authors' Contributions

All authors contributed to the conception and design of the study. EH and IK were responsible for data collection. EH, JP, and AV contributed to data processing. EH and AV performed data analyses. AV wrote the first draft of the article. All authors revised the draft and approved the final version of the manuscript to be submitted.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 14 KB - jmir_v23i6e25529_app1.docx]

References

1. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. Circulation 2014 Jun 24;129(25 Suppl 2):S102-S138 [FREE Full text] [doi: 10.1161/01.cir.0000437739.71477.ee] [Medline: 24222017]
2. Shieh C, Knisely MR, Clark D, Carpenter JS. Self-weighing in weight management interventions: A systematic review of literature. Obes Res Clin Pract 2016 Feb 17;10(5):493-519. [doi: 10.1016/j.orcp.2016.01.004] [Medline: 26896865]
3. 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]

4. 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]
5. Madigan CD, Daley AJ, Lewis AL, Aveyard P, Jolly K. Is self-weighing an effective tool for weight loss: a systematic literature review and meta-analysis. Int J Behav Nutr Phys Act 2015 Aug 21;12:104. [doi: 10.1186/s12966-015-0267-4] [Medline: 26293454]
6. Rosenbaum DL, Espel HM, Butryn ML, Zhang F, Lowe MR. Daily self-weighing and weight gain prevention: a longitudinal study of college-aged women. J Behav Med 2017 Oct;40(5):846-853. [doi: 10.1007/s10865-017-9870-y] [Medline: 28689248]
7. Bertz F, Pacanowski CR, Levitsky DA. Frequent Self-Weighing with Electronic Graphic Feedback to Prevent Age-Related Weight Gain in Young Adults. Obesity (Silver Spring) 2015 Oct;23(10):2009-2014. [doi: 10.1002/oby.21211] [Medline: 26414563]
8. Helander EE, Vuorinen AL, Wansink B, Korhonen IKJ. Are breaks in daily self-weighing associated with weight gain? PLoS One 2014;9(11):e113164. [doi: 10.1371/journal.pone.0113164] [Medline: 25397613]
9. Linde JA, Jeffery RW, French SA, Pronk NP, Boyle RG. Self-weighing in weight gain prevention and weight loss trials. Ann Behav Med 2005 Dec;30(3):210-216. [doi: 10.1207/s15324796abm3003_5] [Medline: 16336072]
10. VanWormer JJ, Martinez AM, Martinson BC, Crain AL, Benson GA, Cosentino DL, et al. Self-weighing promotes weight loss for obese adults. Am J Prev Med 2009 Jan;36(1):70-73. [doi: 10.1016/j.amepre.2008.09.022] [Medline: 18976879]
11. Steinberg DM, Bennett GG, Askew S, Tate DF. Weighing every day matters: daily weighing improves weight loss and adoption of weight control behaviors. J Acad Nutr Diet 2015 Apr;115(4):511-518 [FREE Full text] [doi: 10.1016/j.jand.2014.12.011] [Medline: 25683820]
12. VanWormer JJ, Linde JA, Harnack LJ, Stovitz SD, Jeffery RW. Self-weighing frequency is associated with weight gain prevention over 2 years among working adults. Int J Behav Med 2012 Sep;19(3):351-358 [FREE Full text] [doi: 10.1007/s12529-011-9178-1] [Medline: 21732212]
13. LaRose JG, Lanoye A, Tate DF, Wing RR. Frequency of self-weighing and weight loss outcomes within a brief lifestyle intervention targeting emerging adults. Obes Sci Pract 2016 Mar;2(1):88-92 [FREE Full text] [doi: 10.1002/osp4.24] [Medline: 27668087]
14. Gokee LaRose J, Tate DF, Gorin AA, Wing RR. Preventing weight gain in young adults: a randomized controlled pilot study. Am J Prev Med 2010 Jul;39(1):63-68 [FREE Full text] [doi: 10.1016/j.amepre.2010.03.011] [Medline: 20537843]
15. Houston M, vanDellen M, Cooper JA. Self-weighing Frequency and Its Relationship with Health Measures. Am J Health Behav 2019 Sep 01;43(5):975-993. [doi: 10.5993/AJHB.43.5.9] [Medline: 31439103]
16. Crane MM, Gavin K, Wolfson J, Linde JA. How Accurate are Recalls of Self-Weighing Frequency? Data from a 24-Month Randomized Trial. Obesity (Silver Spring) 2018 Aug;26(8):1296-1302 [FREE Full text] [doi: 10.1002/oby.22239] [Medline: 30070045]
17. Willett W. Nutritional epidemiology, 3rd ed. Oxford, UK: Oxford University Press; 2012.
18. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. The efficacy of a daily self-weighing weight loss intervention using smart scales and e-mail. Obesity (Silver Spring) 2013 Sep;21(9):1789-1797 [FREE Full text] [doi: 10.1002/oby.20396] [Medline: 23512320]
19. Zheng Y, Sereika SM, Burke LE, Olgin JE, Marcus GM, Aschbacher K, et al. Temporal patterns of self-weighing behavior and weight changes assessed by consumer purchased scales in the Health eHeart Study. J Behav Med 2019 Oct;42(5):873-882 [FREE Full text] [doi: 10.1007/s10865-018-00006-z] [Medline: 30649648]
20. 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]
21. Sperrin M, Rushton H, Dixon WG, Normand A, Villard J, Chieh A, et al. Who Self-Weighs and What Do They Gain From It? A Retrospective Comparison Between Smart Scale Users and the General Population in England. J Med Internet Res 2016 Jan 21;18(1):e17 [FREE Full text] [doi: 10.2196/jmir.4767] [Medline: 26794900]
22. Vivanti A, Yu L, Palmer M, Dakin L, Sun J, Campbell K. Short-term body weight fluctuations in older well-hydrated hospitalised patients. J Hum Nutr Diet 2013 Oct;26(5):429-435. [doi: 10.1111/jhn.12034] [Medline: 23521346]
23. Racette SB, Weiss EP, Schechtman KB, Steger-May K, Villareal DT, Obert KA, et al. Influence of weekend lifestyle patterns on body weight. Obesity (Silver Spring) 2008 Aug;16(8):1826-1830. [doi: 10.1038/oby.2008.320] [Medline: 18551108]
24. Orsama AL, Mattila E, Ermes M, van Gils M, Wansink B, Korhonen I. Weight rhythms: weight increases during weekends and decreases during weekdays. Obes Facts 2014;7(1):36-47 [FREE Full text] [doi: 10.1159/000356147] [Medline: 24504358]
25. Frie K, Hartmann-Boyce J, Jebb S, Oke J, Aveyard P. Patterns in Weight and Physical Activity Tracking Data Preceding a Stop in Weight Monitoring: Observational Analysis. J Med Internet Res 2020 Mar 17;22(3):e15790 [FREE Full text] [doi: 10.2196/15790] [Medline: 32181749]
26. Privacy Policy for Withings Products and Services. Withings. URL: https://www.withings.com/fi/en/legal/privacy-policy [accessed 2015-01-01]

27. Turicchi J, O'Driscoll R, Horgan G, Duarte C, Palmeira AL, Larsen SC, et al. Weekly, seasonal and holiday body weight fluctuation patterns among individuals engaged in a European multi-centre behavioural weight loss maintenance intervention. PLoS One 2020;15(4):e0232152 [FREE Full text] [doi: 10.1371/journal.pone.0232152] [Medline: 32353079]
28. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2021-06-11]
29. Wickham H, François R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. R Package Version 1.0.2 2020.
30. Pinheiro J, Bates D, DebRoy S, Sarkar D. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1.-151 2020 [FREE Full text]
31. Brockmann AN, Eastman A, Ross KM. Frequency and Consistency of Self-Weighing to Promote Weight-Loss Maintenance. Obesity (Silver Spring) 2020 Jul;28(7):1215-1218 [FREE Full text] [doi: 10.1002/oby.22828] [Medline: 32437055]
32. Fahey MC, Klesges RC, Kocak M, Wang J, Talcott GW, Krukowski RA. Do the holidays impact weight and self-weighing behaviour among adults engaged in a behavioural weight loss intervention? Obes Res Clin Pract 2019;13(4):395-397 [FREE Full text] [doi: 10.1016/j.orcp.2019.05.001] [Medline: 31182293]
33. Pacanowski CR, Bertz FC, Levitsky DA. Daily Self-Weighing to Control Body Weight in Adults: A Critical Review of the Literature. Sage Open 2014;4(4):1-16 [FREE Full text] [doi: 10.1177/2158244014556992] [Medline: 27127719]
34. Tanenbaum ML, Ross KM, Wing RR. Overeat today, skip the scale tomorrow: An examination of caloric intake predicting nonadherence to daily self-weighing. Obesity (Silver Spring) 2016 Nov;24(11):2341-2343 [FREE Full text] [doi: 10.1002/oby.21650] [Medline: 27619935]

Abbreviations

HeH: Heart eHealth study
LME: linear mixed effect

Edited by R Kukafka, G Eysenbach; submitted 20.11.20; peer-reviewed by M Sperrin; comments to author 11.12.20; revised version received 20.12.20; accepted 17.05.21; published 28.06.21.

Please cite as:
Vuorinen AL, Helander E, Pietilä J, Korhonen I
Frequency of Self-Weighing and Weight Change: Cohort Study With 10,000 Smart Scale Users
J Med Internet Res 2021;23(6):e25529
URL: https://www.jmir.org/2021/6/e25529
doi:10.2196/25529
PMID:34075879

Original Paper

Needs and Preferences of Middle-Aged and Older Adults in Taiwan for Companion Robots and Pets: Survey Study

Ching-Ju Chiu[1*], PhD; Shiuan Hsieh[1*], MS; Chia-Wei Li[1], MS

Institute of Gerontology, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[*] these authors contributed equally

Corresponding Author:
Ching-Ju Chiu, PhD
Institute of Gerontology, College of Medicine
National Cheng Kung University
No. 1, University Road
Tainan, 70101
Taiwan
Phone: 886 6 2353535 ext 5739
Fax: 886 6 3028175
Email: cjchiu@mail.ncku.edu.tw

Abstract

Background: In recent years, robots have been considered a new tech industry that can be used to solve the shortage in human resources in the field of health care. Also, animal-assisted therapy has been used to provide assistance, companionship, and interaction among the elderly and has been shown to have a positive impact on their emotional and psychological well-being. Both pets and robots can provide dynamic communication and positive interaction patterns. However, preferences for middle-aged and older adults in this regard are not clear.

Objective: This study explored the degree of acceptance of robots and pets as partners in later life and to determine the needs and preferences of elderly individuals related to companion robots.

Methods: A total of 273 middle-aged and older adults aged ≥45 years and living in the community were invited to answer a structured questionnaire after watching a companion robot video. Sociodemographic data, physical health status and activities, experience with technology, eHealth literacy, and acceptance and attitude toward robots and pets were recorded and analyzed using multinomial logistic regression analysis.

Results: Age, level of education, type of dwelling, occupation, retirement status, number of comorbidities, experience with pets, experience using apps, and eHealth literacy were significantly associated with acceptance of robots and pets. Middle-aged and older women preferred robots with an animal-like appearance, while men preferred robots that resembled a human adult. In terms of robot functions, participants preferred a companion robot with dancing, singing, storytelling, or news-reporting functions. Participants' marital status and whether or not they lived alone affected their preference of functions in the companion robot.

Conclusions: Findings from this study inform the development of social robots with regard to their appearance and functions to address loneliness in later life in fast-aging societies.

(J Med Internet Res 2021;23(6):e23471) doi:10.2196/23471

KEYWORDS
middle-aged adults; older adults; companionship demand; robot; pet; acceptance

Introduction

Along with increases in the size of the aging population, the demands for care and medical and health care manpower for the elderly population are also increasing. Determining how to adapt to these changes, using limited resources to meet the needs of care recipients, and reducing the burden on caregivers so that middle-aged and older adults have a high quality of life in their old age is an important issue that cannot be ignored.

Cowan [1] divided the issues to be faced by an aging society into 8 categories: dependent living, fall risk, chronic disease, dementia, social isolation, depression, poor well-being, and poor medication management. However, an existing literature review pointed out that in order to establish relevant advanced-age health technologies designed to solve the issues mentioned above, the issues could be divided into 6 groups: general information and communications technology (ICT), robotics, telemedicine, sensor technology, medication management applications, and video games [1].

Over the past decade, the elderly population has been the demographic with the fastest growing use of technological products such as the internet and computers [2,3], and a growing number of studies have shown that health-related ICTs can effectively reduce medical expenditures and care costs and enhance the quality of life of middle-aged and older adults [4,5]. In addition, technological products can help middle-aged and older adults live independently at home and provide health care and medical services in remote areas through mobile health (mHealth) strategies [6]; among these technological products are robots that can assist humans in performing repetitive and dangerous work and become the additional manpower needed for health care [7].

A health care robot is a robot that monitors or promotes physical and mental health and mitigates social psychological problems in the elderly. According to their functions, these robots can be divided into 2 types: rehabilitation robots and social robots [8]. Rehabilitation robots are auxiliary devices that provide physical assistance and make it easier for users to perform physical tasks. They include such things as smart wheelchairs, artificial limbs, and exoskeletons. Social robots interact with the elderly, providing companionship or improving daily life. These robots can be further divided into service-type robots and companionship robots. The function of the service-type robot is mainly related to supporting the independent life of the elderly individual, such as assisting with eating, bathing, toileting, or dressing, as well as performing housework and providing health and safety monitoring. A companionship robot promotes the physical and mental health of elderly persons and enhances their quality of life through companionship, such as the robotic seal PARO that accompanies elderly individuals with dementia; the robot Huggable, which was specially developed for elderly care experimental research; and the robotic dog Aibo, which was intended to improve the quality of life of older individuals and disabled patients [9,10]. Studies have shown that older people prefer less human-looking robots [11,12] and especially enjoy pet-like robots, which are widely used in the care of elderly persons with intellectual disabilities and provide pet-like companionship in lieu of real animals [13,14]. For example, the therapeutic robotic seal pet PARO, which was developed in Japan in 2004, has a body covered with more than 100 sensors and can interact with people. Survey results show that because PARO's appearance is unfamiliar to people, it is less likely that people will feel a sense of artificial interaction with an animal, and it is more likely to be accepted by the elderly [15]. Many studies have also shown that PARO can improve depression, increase social interaction, and positively stimulate cognitive functions in elderly persons with dementia [16,17], which suggests that robot-assisted therapy is a new therapeutic tool for use among the elderly [18-20]. According to the literature review [21], robot-assisted therapy is beneficial to the moods and behavior of elderly persons. A pet-like social robot can stimulate elderly persons to interact and talk with others and remind older adults of their past experiences with companion pets, while posing fewer concerns about safety (such as attacks or bites) and hygiene (allergies, infections, or dirt) that are associated with real pets. Older adults with dementia can also get the same emotional comfort from robot-assisted therapy as they would from their interaction with real pets. As a result, a pet-like robot provides not only simple entertainment but also assistance, companionship, therapy, interaction, and stimulation, as well as other functions and services [21].

However, technology may not be a substitute for human assistance, companionship, and interaction. A study was conducted to enable "robotic dog doctors" to accompany the elderly through animal-assisted therapy, and the results of the study showed that it had a positive effect on the mental and social health of the elderly participants. The study indicated that animal-assisted therapy can improve emotional and behavioral problems, as well as problems with aggression, in elderly individuals with dementia and can have a positive effect on the mental and social health of all elderly persons. Animal-assisted therapy is often recommended as a goal-oriented nonpharmaceutical therapy for mental problems [22]. For example, a study by Garrity et al [23] on widowed and socially isolated elderly persons over the past year found that those who had no experience with keeping pets were more depressed than those who had such experience. There are also studies showing that keeping pets is related to the survival rate of cardiovascular disease in the elderly [24], suggesting that pet companionship has a curative effect that cannot be ignored in clinical care and treatment. Some scholars have suggested that patients with dementia can experience a reduction in their degree of loneliness and engage in social interaction by interacting with robot pets and get pleasure and attention from it as well as spiritual comfort [25,26]. Thus, robot pets provide a new therapeutic option for the elderly with dementia. Furthermore, animal protection regulations in countries in addition to a lack of adequate animal training makes robot pets more attractive than animal pets. For example, people generally have doubts about the safety and health of animals in Taiwan, which leads to a lot of restrictions on their implementation in therapy [27]. Therefore, robots or robot pets provide the elderly with dynamic, 2-way communication and a positive interaction mode, which can be regarded as another option for them in later life. Robots can do more dangerous and tiring work in the home care of elderly persons, but they may undergo failure and present financial and ethical issues. Although pets have more spontaneous reactions and richer emotional responses and can provide more tactile stimulation, they present safety and health issues that must be taken into consideration, as well as extra time-consuming care requirements. The aforementioned factors affect the user's choice. According to the theory of planned behavior proposed by Ajzen [28], the occurrence of a behavior depends on the intensity of people's intentions, and the intensity of intentions is determined by 3 factors: attitude, subjective norms, and perceived behavioral control. Therefore, it is important to explore what factors affect the acceptance of robots and pets among middle-aged and older adults in Taiwan and to understand whether these factors correspond to the theory of planned behavior.

Thodberg et al [29] performed a study that compared pets with robotic dogs. At the beginning of the study, the robotic pet PARO and real pets (dogs) had the same impact on residents. However, with increases in interaction time, residents decreased their conversation and eye contact with PARO, but their focus on and interaction with the dogs remained stable [29]. The study also found that the real animals had more spontaneous and richer emotional responses than the robotic pets and that subjects could get more active tactile stimulation. Compared with toy animals, both robotic pets and real animals can provide 2-way dynamic communication, so it is feasible to use a robot/robot pet as a companion object and an auxiliary technological device for stimulating the sensory and cognitive functions of elderly individuals. However, most of the existing studies exploring the effectiveness of robot pet interventions in the elderly population (eg, psychological and behavioral effects and impact on quality of life) have been conducted on institutionalized elderly individuals with dementia for which long-term care was provided [30-33]; few of the studies have included elderly persons in the community as the study population. In particular, there has been a lack of study on the attitudes, degree of acceptance, and needs of middle-aged and older adults as they relate to robots. In the past, there have also been no studies comparing pets with robots in terms of their use as companion objects. This study is aimed toward closing these gaps in the existing literature to discuss the companion needs of middle-aged and older adults in Taiwan in order to understand their choices of robots or live pets as companion objects in later life and to further discuss the preferences of middle-aged and older adults for companionship robots, as well as other related factors.

Methods

Participants

Adults over age 45 years in Taiwan were invited to participate in this study using a convenience sampling method based on the sample selection standard. The number of participants needed for the study was determined as the number of variables (n=26) multiplied by 10. Thus, a total of 273 older adults living in the community comprised the sample. The questionnaire was distributed in gathering places for the elderly in Taiwan such as community universities, senior citizen learning centers, community care strongholds, day-care strongholds, and hospital clinics all over Taiwan. The inclusion criteria for participants were as follows: (1) able to communicate in Mandarin and Taiwanese; (2) willing to be interviewed by researchers, to fill out the questionnaire on their own, or to fill out an electronic questionnaire with a tablet computer; and (3) agreed to participate in the study and to sign a consent form. The exclusion conditions were as follows: (1) resided somewhere other than Taiwan, and (2) were suffering from moderate to severe cognitive impairment or unable to answer questions without coercion.

Measures

A structural questionnaire was used as the research tool. The content of the questionnaire included 4 parts: sociodemographic data, physical and mental health status and activities, technology use and eHealth literacy, and robot and pet experience.

The sociodemographic data included age, gender, education level, marital status, number of children, place of residence, whether living alone or not, working status, economic status, self-rated health status, and number of chronic diseases.

The physical and mental health status and activities included social participation, leisure activities, social support, depression status, and personality traits. Among them, social participation and leisure activities were measured using the Ministry of Health and Welfare Taiwan Longitudinal Study on Aging questionnaire [34]. Social support was measured using the Inventory of Socially Supportive Behavior (ISSB [35,36]). This inventory consists of 10 questions relating to 4 types of social support: emotional support (3 questions), information support (2 questions), substantive support (2 questions), and social integration (3 questions). In the scoring system for the ISSB, a score of 1 represents unsatisfied, 2 represents neutral, and 3 represents satisfied; the higher the score, the higher the perceived social support. An overall internal consistency coefficient of .91 represents emotional support for each support type α reliability coefficient; a coefficient of .81 represents information support; a coefficient of .73 represents substantive support; and a coefficient of .81 represents social integration [37]. The depression status was measured using the simplified, 10-item version of the Center for Epidemiologic Studies Depression Scale (CES-D [38]), which was translated into Chinese. The CES-D comprises 10 positive and 10 negative questions that are scored on Likert scales ranging from 0 to 3, with the total score ranging from 0 to 30 and a total score of greater than 10 representing depression. The overall internal consistency Cronbach α value ranges from .78 to .87 [39]. The personality assessment was carried out using the International Personality Item Pool (IPIP) big 5 personality scale developed by Goldberg in 1992 [40], which was first translated into a simplified Chinese version [41] and then changed to a traditional Chinese version with customary modifiers used by the Taiwanese population [42]. A single factor was screened out from the original 50 questions, and the questions with higher factor loadings in each domain were developed into a new, 15-item version of the IPIP (IPIP-15). This simplified version of the IPIP big 5 personality scale is divided into 5 dimensions—extroversion, friendliness, rigorousness, emotional stability, and intelligence/imagination—that are scored from 0 to 5, where 1=imprecise, 2=slightly imprecise, 3=ordinary, 4=slightly precise, and 5=very precise. The Cronbach α reliability coefficient judging the internal consistency of each scale ranged between .67 and .83, and the factor loading ranged between .61 and .83, indicating convergent validity. The correlation between the IPIP-15 and the personality scales corresponding to the original IPIP-50 ranged between .81 and .88, which indicated that the convergent validity was acceptable [42].

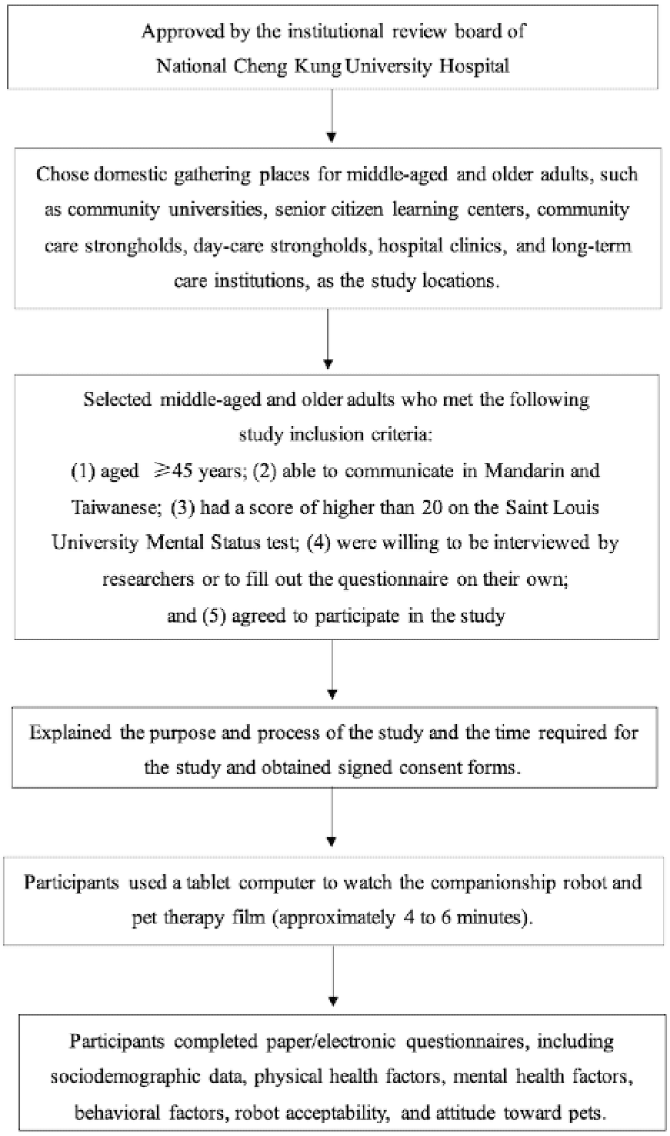
Participants' experience in the use of technology, networking, and eHealth literacy were also assessed. The question about experience with the use of technology and networking was answered by subjects based on their past experience (ie, whether they had experience using the internet and downloading and using mobile apps). The question about eHealth literacy was assessed using the eHealth Literacy Scale (eHEALS), which is an 8-item measure that assesses the participant's internet use and search skills, ability to evaluate online content, and confidence in their internet-searching abilities. This scale is scored with 4 points, with the options being entirely disagree, disagree, agree, and strongly agree, which are scored 1, 2, 3, and 4, respectively. The internal consistency Cronbach α reliability coefficient for each item is .88, and the factor load ranges between .60 and .84 [43,44].

Experience with and acceptance of robots and pets included the acceptability of choosing a robot or pet as a companion object in later life, the type of companionship robot desired/favored, and past experience with keeping pets. Among these indices, the question about the acceptability of robots/pets was answered based on the response, "Acceptability of choosing a robot or pet as a companion object in later life." It was scored from 0 to 10, with 0 representing completely unacceptable and 10 representing quite acceptable, and the higher the score, the higher the degree of acceptability. In addition, when the type of companion robots desired/favored was also evaluated, subjects were requested to choose the types and functions of the robots according to their preferences or needs. The choices included the companionship robot's services (eg, assisting with family tasks, health monitoring, safety monitoring), skills (eg, juggling, dancing, singing), interaction (eg, chatting, storytelling, news reporting, joke telling, providing child-like dialogue), expression, and appearance (eg, resembling an animal, human infant or adult, or nonbiological form), as well as other functions.

Procedure

The study was conducted between May and June 2018. A cross-sectional survey research method was used to survey the degree of acceptance and factors related to the choices made by middle-aged and older adults in Taiwan of a robot or pet as their companion object in later life. This study was approved by the institutional review board (IRB) of National Cheng Kung University Hospital in Taiwan (No. A-ER-105-509). The study collection methods and procedures are shown in Figure 1.

Figure 1. Research and data collection flowchart.



Approved by the institutional review board of National Cheng Kung University Hospital

Chose domestic gathering places for middle-aged and older adults, such as community universities, senior citizen learning centers, community care strongholds, day-care strongholds, hospital clinics, and long-term care institutions, as the study locations.

Selected middle-aged and older adults who met the following study inclusion criteria:
(1) aged ≥45 years; (2) able to communicate in Mandarin and Taiwanese; (3) had a score of higher than 20 on the Saint Louis University Mental Status test; (4) were willing to be interviewed by researchers or to fill out the questionnaire on their own; and (5) agreed to participate in the study

Explained the purpose and process of the study and the time required for the study and obtained signed consent forms.

Participants used a tablet computer to watch the companionship robot and pet therapy film (approximately 4 to 6 minutes).

Participants completed paper/electronic questionnaires, including sociodemographic data, physical health factors, mental health factors, behavioral factors, robot acceptability, and attitude toward pets.

With the consent of the IRB of National Cheng Kung University Hospital in Taiwan, middle-aged and older adults in Taiwan were invited to participate in the study, and gathering places for the middle-aged and older adults, such as community universities, senior citizen learning centers, community care strongholds, and day-care strongholds, were chosen as places to distribute questionnaires. Middle-aged and older adults who met the following inclusion criteria were selected: (1) aged ≥45 years, (2) able to communicate in Mandarin and Taiwanese, (3) had a Saint Louis University Mental Status test score higher than 20, (4) willing to be interviewed by the researchers or to fill out the questionnaire on their own, and (5) agreed to participate in the study and sign a consent form. The researchers explained to the subjects the study's purpose, process, and duration. After signing the consent form, the subjects used a tablet computer to watch a companionship robot film, after which they completed a questionnaire.

A companionship robot film was used in the study to provide a brief introduction to each type of companionship robot and was explained by the researchers at the time it was played. The film included robots with various physical characteristics (Multimedia Appendix 1), such as resembling an adult, an infant, an animal, or a nonbiological object, and presented the different functions of each robot, such as assisting with family tasks, health monitoring, safety monitoring, and other services, and the content of the robots' interactions with users, such as chatting, reporting news, reporting weather, singing, dancing, and making various expressions, among other functions.

Statistics

The descriptive statistics included an analysis of the sociodemographic variables, physical health factors, past experience with keeping pets, and experience in the use of technology as they related to variables such as the acceptability of choosing a robot or pet as a companion object in later life using a *t* test and a 1-way analysis of variance. Differences among the sociodemographic variables, physical health factors, past experience with keeping pets, and experience with the use of technology, as well as other variables related to the elderly, were verified using the chi-square test among 4 groups of robot/pet preferences. The analysis of the correlation between the continuous variables was related to the Pearson correlation and included age, number of years living alone, IPIP-15 score, CES-D score, ISSB score, eHEALS score, and the level of acceptance of a robot or pet as a companion object in later life. A multiple regression analysis was used to analyze the between-variable correlations, such as gender, age, education level, living alone or not, retirement status, number of comorbidities, ISSB score, eHEALS score, IPIP-15 score, and acceptance of either robots or pets. Finally, a multinomial logistic regression analysis was used to analyze the predictive power of different groups of robot or pet acceptability in the middle-aged and older adults based on the following variables: (1) both robots and pets were highly acceptable (HH), (2) preferred choice was a robot (HL), (3) preferred choice was a pet (LH), and (4) neither robots nor pets were acceptable (LL).

Results

Descriptive Analysis of the Subjects' Basic Data

For the purpose of discussing the degree to which the middle-aged and older adults preferred a robot or pet as their companion object in later life in Taiwan, the questionnaires were distributed at 6 community care strongholds, 5 community centers, and 3 large-scale activities related to respecting the elderly and Mother's Day events in the north, middle, and southern parts of Taiwan. A total of 273 subjects who met the inclusion criteria were selected out of 300 middle-aged and older adults living in the community who were aged ≥45 years, and a total of 240 valid questionnaires were obtained after those with missing data or incorrect answers (n=33) were removed. The minimum and maximum ages of the participants were 45 years and 94 years, respectively. The average age was 60.68 years, and there were 172 (71.7%) female participants and 68 (28.3%) male participants. The majority of participants were highly educated (183/240, 76.2%), had a partner (170/240, 70.8%), lived in the city (214/240, 89.2%), did not live alone (215/240, 89.6%), lived with children (221/240, 92.1%), had no experience with using robots (197/240, 82.1%), had experience using the internet (196/240, 81.7%), could download and use an app (184/240, 76.7%), had experience with keeping pets (152/240, 63.6%), and had no experience with animal-assisted therapy (226/240, 94.2%). The subjects reported an average of 0.65 chronic conditions and an average self-reported health score of 3.53 (out of 5). On average, subjects' level of acceptance of robots and pets was 5.69 points and 4.72 points, respectively. A detailed chart of the data distribution is shown in Table 1.

Table 1. Basic sociodemographic data of participants.

Characteristic		Preference of robot or pet as a companion				
	Full sample (N=240)	Both are highly acceptable[a] (n=81, 33.8%)	Prefer robot (n=42, 17.5%)	Prefer pet (n=56, 23.3%)	Neither is acceptable (n=61, 25.4%)	Verification value (F/χ^2)
Age (years), mean (SD)	60.68 (10.496)					13.65
45-54	85 (35.4)	37 (45.7)	8 (19.0)	22 (39.3)	18 (29.5)	
55-64	73 (30.4)	21 (25.9)	13 (31.0)	20 (35.7)	19 (31.1)	
65-74	53 (22.1)	14 (17.7)	14 (33.3)	10 (17.9)	15 (24.6)	
≥75	29 (12.1)	9 (11.1)	7 (16.7)	4 (7.1)	9 (14.8)	
Gender, n (%)						1.34
Male	68 (28.3)	22 (27.2)	12 (28.6)	19 (33.9)	15 (24.6)	
Female	172 (71.7)	59 (72.8)	30 (71.4)	37 (66.1)	46 (75.4)	
Education level, n (%)						9.41
Below primary school	57 (23.8)	15 (18.5)	15 (35.7)	11 (19.6)	16 (26.2)	
Secondary school/senior high school (higher vocational school)	73 (30.4)	23 (28.4)	8 (19.0)	19 (33.9)	23 (37.7)	
University and above	110 (45.8)	43 (53.1)	19 (45.2)	26 (46.4)	22 (36.1)	
Marital status, n (%)						4.48
Unmarried/widowed/no partner	70 (29.2)	18 (22.2)	16 (38.1)	15 (26.8)	21 (34.4)	
Married or has a partner	170 (70.8)	63 (77.8)	26 (61.9)	41 (73.2)	40 (65.6)	
Residence, n (%)						0.866[b]
City	214 (89.2)	71 (33.2)	38 (90.5)	49 (87.5)	56 (91.8)	
Village	26 (10.8)	10 (12.4)	4 (9.5)	7 (12.5)	5 (8.2)	
Type of dwelling, n (%)						10.19
House	109 (45.4)	42 (51.9)	20 (47.6)	23 (41.1)	24 (39.3)	
Apartment building without elevator	39 (16.3)	13 (16.0)	4 (9.5)	6 (10.7)	16 (26.2)	
Apartment building with elevator	92 (38.3)	26 (32.1)	18 (42.9)	27 (48.2)	21 (34.4)	
Lives alone, n (%)						4.244[b]
Yes	25 (10.4)	7 (8.6)	5 (11.9)	3 (5.4)	10 (16.4)	
No	215 (89.6)	74 (91.4)	37 (88.1)	53 (94.6)	51 (83.6)	
Number of children, mean (SD)	2.16 (1.117)					
Number of years living alone (years), mean (SD)	0.90 (3.491)	0.69 (3.204)	1.55 (4.910)	0.73 (3.419)	0.87 (2.699)	0.62 (1.146)[b]
Lives with children, n (%)						1.146[b]
Yes	221 (92.1)	75 (92.6)	37 (88.1)	52 (92.9)	57 (93.4)	
No	19 (7.9)	6 (7.4)	5 (11.9)	4 (7.1)	4 (6.6)	
Type of occupation, n (%)						10.05
Unskilled	76 (31.7)	20 (24.7)	16 (38.1)	14 (25.0)	26 (42.6)	
Semiskilled or skilled	87 (36.3)	32 (39.5)	12 (28.6)	20 (35.7)	23 (37.7)	
Professional/senior managers	77 (32.1)	29 (35.8)	14 (33.3)	22 (39.3)	12 (19.7)	
Retirement status, n (%)						9.40*
Retired	118 (49.2)	32 (39.5)	27 (64.3)	24 (42.9)	35 (57.4)	
Employed	122 (50.8)	49 (60.5)	15 (35.7)	32 (57.1)	26 (42.6)	
Number of chronic diseases, mean (SD)	0.65 (0.878)	0.43 (0.670)	1.02 (1.047)	0.70 (0.807)	0.66 (0.981)	4.434**

Characteristic		Preference of robot or pet as a companion				
	Full sample (N=240)	Both are highly acceptable[a] (n=81, 33.8%)	Prefer robot (n=42, 17.5%)	Prefer pet (n=56, 23.3%)	Neither is acceptable (n=61, 25.4%)	Verification value (F/χ^2)
Self-rated health (1 to 5 points), mean (SD)	3.53 (0.731)	3.58 (0.756)	3.50 (0.672)	3.43 (0.828)	3.59 (0.642)	0.644
Self-rated financial status (1 to 5 points), mean (SD)	3.83 (0.062)	3.88 (0.509)	3.81 (0.594)	3.75 (0.745)	3.85 (0.654)	0.501
Has experience using robots, n (%)						3.16
Yes	43 (17.9)	18 (22.2)	7 (16.7)	6 (10.7)	12 (19.7)	
No	197 (82.1)	63 (77.8)	35 (83.3)	50 (89.3)	49 (80.3)	
Has experience using the internet, n (%)						7.80
Yes	196 (81.7)	65 (80.2)	30 (71.4)	52 (92.9)	49 (80.3)	
No	44 (18.3)	16 (19.8)	12 (28.6)	4 (7.1)	12 (19.7)	
Experience using apps, n (%)						2.51
No experience with using apps/not able to download	56 (23.3)	20 (35.7)	13 (23.2)	12 (21.4)	11 (19.6)	
Can download and use apps	184 (76.7)	61 (75.3)	29 (69.0)	44 (78.6)	50 (82.0)	
Has experience keeping pets, n (%)						20.55***
Yes	152 (63.3)	63 (77.8)	18 (42.9)	40 (71.4)	31 (50.8)	
No	88 (36.7)	18 (22.2)	24 (57.1)	16 (28.6)	30 (49.2)	
Animal-assisted therapy experience, n (%)						1.806[b]
Has no animal care experience	226 (94.2)	76 (93.8)	40 (95.2)	51 (91.1)	59 (96.7)	
Has animal care experience	14 (5.8)	5 (6.2)	2 (4.8)	5 (8.9)	2 (3.3)	
IPIP-15[c] >score, mean (SD)						
Extroversion	11.51 (2.626)	11.69 (2.391)	11.55 (2.487)	11.50 (2.683)	11.25 (2.987)	0.335
Friendliness	11.78 (2.141)	11.77 (2.260)	11.76 (2.034)	12.02 (2.244)	11.59 (1.978)	0.390
Rigorousness	11.90 (2.405)	11.57 (2.617)	12.19 (2.287)	12.29 (2.078)	11.80 (2.455)	1.235
Emotional stability	11.28 (2.746)	10.91 (2.651)	11.71 (3.263)	11.34 (2.345)	11.41 (2.831)	0.884
Intelligence/imagination	9.72 (2.604)	9.81 (2.393)	9.43 (2.881)	10.09 (2.678)	9.44 (2.617)	0.815
CES-D[d] score, mean (SD)	5.93 (5.330)	6.07 (5.422)	6.10 (5.938)	5.79 (5.098)	5.74 (5.092)	0.072
Social participation, mean (SD)	2.25 (2.124)	2.14 (2.223)	2.81 (2.211)	2.05 (2.211)	2.18 (1.812)	1.234
Leisure activities, mean (SD)	7.85 (2.147)	8.01 (2.009)	7.93 (2.005)	7.75 (2.250)	7.69 (2.349)	0.325
Social support, mean (SD)	24.88 (4.462)	24.67 (4.693)	24.88 (4.743)	25.05 (4.020)	25.00 (4.431)	0.104
eHealth Literacy Scale score, mean (SD)	21.65 (7.773)	22.04 (7.825)	21.36 (8.316)	23.43 (6.494)	19.69 (8.123)	2.400
Acceptability[a] (0 to 10 points), mean (SD)						
Robot	5.69 (3.142)					
Pet	4.72 (3.564)					

[a] Acceptability was deemed to be high or low if it was higher or lower, respectively, than the average points of acceptability.

[b] Because the number of people who had few expectations was less than 5, the test was conducted using Fisher exact test, and the results showed no significant differences.

[c] IPIP-15: 15-item International Personality Item Pool big 5 personality scale.

[d] CES-D: Center for Epidemiologic Studies Depression Scale.

**P*<.05; ***P*<.01; and ****P*<.001.

Factors Associated With Intention to Use a Robot or Pet as Their Companion Object in Later Life

The variables related to acceptance of robots and pets were divided into 4 groups: both robots and pets were highly acceptable (HH), preferred a robot (HL), preferred a pet (LH), and neither robots nor pets were acceptable (LL). The level of acceptability was deemed to be high if it was higher than average and was deemed to be low if it was lower than the average. As Table 1 shows, 33.8% (81/240) of the subjects reported both robots and pets to be acceptable (HH), 17.5% (42/240) preferred a robot (HL), 23.3% (56/240) preferred a pet (LH), and 25.4% (61/240) reported neither robots nor pets to be acceptable (LL). The results of the study showed that retirement status (χ^2_3=9.40, *P*=.024), experience with keeping pets (χ^2_3=20.55, *P*=.000), and the number of comorbidities ($F_{3,236}$=4.43, *P*=.005) were all significantly associated with the 4 acceptance groups. However, none of the psychological measures, including personality traits, CES-D score, and social participation, were associated with the preference for robots or pets as their companion.

Results of the multinomial logistic regression analysis used to analyze the variables in terms of their predictive power on the 4 groups of robot/pet acceptability among the middle-aged and older adults is presented in Table 2. The results showed that those with more comorbidities were 1.688 times more likely to fall into the HL group than into the LL group (*P*=.048); those who could download and use an app were 0.170 times more likely to fall into the HL group than the LL group as compared with those who had not used or downloaded apps (*P*=.022). Those who could download and use apps were 0.159 times more likely to fall into the HL group than into the LL group as compared with those who had not used or downloaded apps (*P*=.012). Those with experience with keeping pets were 3.527 times more likely to fall into the HH group than into the LL group as compared with those who had no experience with keeping pets (*P*=.002). Those who had experience with keeping pets were 2.498 times more likely to fall into the LH group than into the LL group as compared with those with no experience in keeping pets (*P*=.034). Those with a higher score on eHEALS were 1.084 times more likely to fall into the HH group than into the LL group (*P*=.039). Those with higher scores on eHEALS were 1.139 times more likely to fall into the HL group than into the LL group (*P*=.005). Finally, those with higher scores on eHEALS were 1.100 times more likely to fall into the LH group than into the LL group (*P*=.020).

Table 2. Multinomial logistic regression analysis of the degree of acceptance of robots and pets by middle-aged and older adults (N=240).[a]

Characteristic	Preference of robot or pet as a companion		
	Both robots and pets are highly acceptable (n=81, 33.8% [OR])	Prefer robots (n=42, 17.5% [OR])	Prefer pets (n=56, 23.3% [OR])
Age, years (reference: ≥75 years old)			
45-54	1.392	0.982	1.849
55-64	0.896	1.953	1.649
65-74	1.066	2.234	1.129
Number of comorbidities	0.877	1.688*	1.348
Lives alone	0.484	0.664	0.275
Male gender	0.863	1.173	1.646
Education level (reference: above university)			
Below primary school	0.939	0.645	1.593
Secondary school/senior high school (higher vocational school)	0.759	0.370	1.080
Retired	0.555	1.015	0.771
Has experience using robots	0.933	0.734	0.358
Has experience using the internet	0.471	0.368	3.350
Can download and use apps	0.271	0.170*	0.159*
Has experience with keeping pets	3.527**	0.784	2.498*
Social support score	0.969	0.953	0.977
eHealth Literacy Scale score	1.084*	1.139**	1.100*

[a] Contrast group: neither robots nor pets are acceptable (n=61; 25.4%).
*$P<.05$; **$P<.01$; and ***$P<.001$.

Subjects' Preferences for Companionship Robot Functions

This section discusses the participants' preferences for the functions of the companionship robot as their companion object in later life. The subjects filled out questionnaires regarding their preferences or needs for functions of the companionship robot. The functions of the companionship robot included (1) family services, (2) health status monitoring, (3) safety monitoring, (4) skill and recreation-type functions (eg, juggling, dancing, singing, storytelling, news reporting, joke telling, the ability to make various expressions), and (5) interactive functions (eg, chatting, providing child-like dialogue).

The results of the study showed that the functions of the companionship robot that the subjects most desired were the skill and recreation-type functions (211/240, 87.9%), followed by family services (185/240, 77.1%), interactive functions (160/240, 66.7%), health status monitoring (147/240, 61.3%), and safety monitoring (144/240, 60.0%), as shown in Table 3. By analyzing the correlation between the sociodemographic characteristics and the function selection for the companionship robot, it was found that in addition to skill and recreation-type functions (60/68, 88.2%) and family services (50/68, 73.5%), male subjects desired health status monitoring functions (46/68, 67.6%) of the companionship robot more than its interactive functions (44/68, 64.7%). Female subjects desired the safety monitoring functions (103/172, 59.9%) of the companionship robot more than its health status monitoring functions (101/172, 58.7%). Among these, although the differences were not statistically significant, the middle-aged and older female subjects still preferred family service–type robots and interactive function–type robots more than the male subjects.

Table 3. Analysis of the preference of middle-aged and older adults for the companionship robot functions (N=240).

Characteristic	Robot functions				
	Skill and recreation-type functions	Family services	Interactive functions	Health status monitoring	Safety monitoring
Full sample, n (%)	211 (87.9)	185 (77.1)	160 (66.7)	147 (61.3)	144 (60.0)
Gender					
Male (n=68), n (%)	60 (88.2)	50 (73.5)	44 (64.7)	46 (67.6)	41 (60.3)
Female (n=172), n (%)	151 (87.8)	135 (78.5)	116 (67.4)	101 (58.7)	103 (59.9)
χ^2	0.009	0.678	0.164	1.636	0.003
Age					
45-54 years (n=85), n (%)	73 (85.9)	74 (87.1)	59 (69.4)	61 (71.8)	60 (70.6)
55-64 years (n=73), n (%)	66 (90.4)	60 (82.2)	46 (63.0)	46 (63.0)	43 (58.9)
≥65 years (n=82)	72 (87.8)	51 (62.2)	55 (67.1)	40 (48.8)	41 (50.0)
χ^2	0.760	16.156***	0.733	9.427**	7.424*
Residential status					
Living alone (n=25), n (%)	23 (92.0)	15 (60.0)	16 (64.0)	15 (60.0)	11 (44.0)
Not living alone (n=215), n (%)	188 (87.4)	170 (79.1)	144 (67.0)	132 (61.4)	133 (61.9)
χ^2	0.438	4.610*	0.089	0.018	2.977
Marital status					
Have a partner (n=170), n (%)	153 (90.0)	134 (78.8)	113 (66.5)	111 (65.3)	107 (62.9)
Have no partner (n=70), n (%)	58 (82.9)	51 (72.9)	47 (67.1)	36 (51.4)	37 (52.9)
χ^2	2.381	0.999	0.010	4.016*	2.101

*P<.05; **P<.01; and ***P<.001.

According to the results of the analysis on residential status and preferences for functions of the companionship robot, 64% (16/25) of the people living alone had a greater preference and demand for interactive functions, which took second place among those living alone in terms of the desired functions of the companionship robot. Among these subjects, whether or not they lived alone was significantly related to the preference for family service (ie, housework) functions (P=.032). The results for choosing skill and recreation-type functions for the companionship robot by those living alone were not statistically significant, but there was still a tendency toward choosing this type of robot.

The results of the analysis on the marital status and the preference for the companionship robot functions showed that for subjects who had no partner, the level of preference and demand for safety monitoring functions was higher than for health status monitoring functions, and having a partner was significantly related to the choice of health status monitoring functions (P=.045).

The subjects were divided into 3 age groups: 45 to 54 years, 55 to 64 years, and ≥65 years. By analyzing the choices made by the elderly in each age group for the functions of the companionship robot, it was found that the subjects ranging in age from 45 to 54 years mainly preferred the family service–type robot followed by the skill and recreation-type robot. However, subjects who were aged 55 to 64 years and over 65 years all chose the skill and recreation-type robot, including the functions of juggling, dancing, singing, storytelling, news reporting, joke telling, and the ability to make various expressions, followed by the family service–type robot. The different ages were significantly related to the choice of family service (*P*<.001), health status monitoring (*P*=.009), and safety monitoring (*P*=.026) functions of the companionship robot.

Subjects' Preferences for the Appearance of the Companionship Robot

This section discusses the preferences of the middle-aged and older adults for the appearance of the companionship robot as their companion object in later life. The subjects filled out a questionnaire according to their preferences or requirements for the appearance of the companionship robot, which included animal, infant, adult, and nonbiological or other form.

It was found that the appearance of the companionship robot that the subjects most desired/preferred was one resembling an animal (94/240, 39.2%), followed by one resembling an adult (72/240, 30.0%), an infant (43/240, 17.9%), and a nonbiological or other form (21/240, 8.9%). The results are shown in Table 4. The analysis of the correlation between the sociodemographic characteristics and the choice of the appearance of the companionship robot showed that male subjects preferred a companionship robot that resembled an adult (29/68, 42.6%), followed by an animal-like appearance (23/68, 33.8%), where gender was found to be significantly related to the preference for the companionship robot to look like a human adult (*P*=.007). The correlations between residential status and marital status with choice of robot appearance did not reach statistical significance, but living alone and with a partner had the same ranking as the full sample in terms of this preference. Regardless of age, the appearance of animals was the most popular choice of robot appearance. The second most preferred robot appearance was that of a human adult among subjects who were aged 45 to 54 years old and those aged 55 to 64 years; subjects who were ≥65 years preferred the infant- and adult-like appearances equally.

Table 4. Analysis of the preference of the middle-aged and older adults for the appearance of the companionship robot (N=240).

Characteristic	Preferred appearance of companionship robot			
	Animal	Adult	Infant	Other
Full sample, n (%)	94 (39.2)	72 (30.0)	43 (17.9)	21 (8.9)
Gender				
Male (n=68), n (%)	23 (33.8)	29 (42.6)	7 (10.3)	4 (5.9)
Female (n=172), n (%)	71 (41.3)	43 (25.0)	36 (20.9)	17 (9.9)
χ^2	1.137	7.227**	3.749	0.977
Age				
45-54 years (n=85), n (%)	36 (42.4)	13 (15.3)	28 (32.9)	8 (9.4)
55-64 years (n=73), n (%)	29 (39.7)	12 (16.4)	26 (35.6)	7 (9.6)
≥65 years (n=82), n (%)	29 (35.4)	18 (22.0)	18 (22.0)	6 (7.3)
χ^2	0.869	3.976	1.414	0.322
Residential status				
Living alone (n=25)	73 (42.9)	51 (30.0)	27 (15.9)	16 (9.4)
Not living alone (n=215)	21 (30.0)	21 (30.0)	16 (22.9)	5 (7.1)
χ^2	3.485	0.000	1.640	0.320
Marital status				
Have a partner (n=170)	12 (48.0)	8 (32.0)	4 (16.0)	1 (4.0)
Have no partner (n=70)	82 (38.1)	64 (29.8)	39 (18.1)	20 (9.3)
χ^2	0.914	0.053	0.070	0.789

***P*<.01

Discussion

Principal Findings

The main purpose of this study was to discuss the level of acceptance of middle-aged and older adults toward a robot or pet as their companion object in later life; to understand the correlation between sociodemographic variables, physical health, mental health, behavioral factors, and preferences for either a robot or pet as a companion object in later life; and to further analyze the needs and preferences of the subjects for the functions and appearance of a companionship robot. The results of the community survey showed that the level of acceptance of subjects in the community toward a pet as their companion object in later life was significantly correlated with their age, with a higher age being associated with a lower average score for acceptance of a pet. Acceptance of pets was significantly correlated with education level, type of occupation, retirement status, number of comorbidities, past experience with keeping pets, an extroverted personality, an intellectual/imaginary personality, and eHealth literacy. In terms of the acceptance toward robots, there were no significant differences in the level of acceptance based on age group. The level of acceptance toward robots was only significantly related to the type of dwelling the subject lived in. No correlation between the acceptance toward robots and other sociodemographic variables, or physical health, mental health, or behavioral factors was observed in this study.

According to the acceptability score, the level of acceptance of the middle-aged and older adults for choosing between a robot or pet as their companion object in later life was divided into 4 groups: both are highly accepted (HH), preferred a robot (HL), preferred a pet (LH), and neither was acceptable (LL). When the average acceptability score was used as the grouping standard, there were significant differences in terms of retirement status, number of comorbidities, and past experience with keeping pets among each group of subjects. The results of the multinomial logistic regression analysis on the key variables showed that the number of comorbidities, experience with keeping pets, experience with using apps, and eHealth literacy had significant predictive power for the level of acceptability among all of the groups; however, gender, intergenerational differences, education level, whether or not subjects lived alone, social support, and past experience with keeping pets had no significant impact on the level of acceptability among the various groups.

According to the theory of planned behavior, attitude, subjective norms, and perceived behavioral control are the 3 factors that determine behavior [28]. In the research, although the reasons for the significant variables were not further explored, we can infer that personality, education level, type of occupation, and past experience with keeping pets, which can affect the experience of life, might be related to personal attitude. The type of dwelling in which a person resides may affect their preference for robots or pets. It may be related to subjective norms; after all, sometimes living conditions such as neighbors or house size might restrict one from keeping pets. On the other hand, age (related to one's functional ability to keep pets), retirement status (ie, how much leisure time someone has), number of comorbidities, experience with using apps, and eHealth literacy might be related to perceived behavioral control. Overall, the results showed that preferences for robots and pets among middle-aged and older adults conformed to the theory of planned behavior.

By analyzing the needs and preferences of middle-aged and older adults for the companionship robot, it was found that these individuals desired/favored robot skills such as juggling, dancing, singing, storytelling, news reporting, joke telling, or the ability to make various expressions, followed by family service, interactive, health status monitoring, and safety monitoring functions. In terms of the appearance of the companionship robot, the middle-aged and older adults preferred the robot to look like an animal, followed by it having an adult-like appearance, infant-like appearance, and nonbiological or other appearance, in that order. By further analyzing the impact of sociodemographic characteristics on the preferences for the companionship robot, it was found that whether the subjects lived alone or not significantly affected the choice of the family service function of the robot, and whether there was a partner or not also significantly affected whether the subjects chose the health status monitoring function of the robot. Male subjects showed a greater preference for an adult-like appearance in the robot than did female subjects, and the difference was statistically significant. Female subjects preferred an animal-like appearance to the robot over an adult-like appearance.

Comparison With Prior Work

Most studies on the appearance of robots have pointed out that elderly individuals prefer less human-looking robots such as pet-like robots, which have been widely used in the past to care for the elderly and are highly accepted by them [12,13,45]. Studies in Japan showed that the robotic seal pet PARO, because its seal-like appearance was unfamiliar to people, did not lead to an unreal sense of interaction with a fake animal and was easily accepted by the elderly [46]. A study discussing robots in the daily lives of the elderly in Taiwan pointed out that older adults were more likely to accept robot pets of traditional pet animals, such as cats or dogs, because the elderly associated the robots with animals they were familiar with, and those who had no experience with keeping pets wanted pet-like robots as pets [47]. A study discussing the needs of the elderly for the companionship robot when they entered the "empty-nest" period indicated that the appearance of a future companionship robot needed to be based on human life experiences. For example, the Hug is a robot that allows the elderly to maintain social and affective interactions by communicating closely with their families. It was designed to look like a human offering a hug and has specific types of communication functions [48]. The SenseChair is a robot that was designed to look like a "chair" that the elderly are familiar with in their daily life [49]. In terms of the robot's functions, the main purpose of using a home-based robot was to obtain the "home service" function, following by providing assistance to people with mobility disabilities, home security management, remote monitoring, emotional pacification, and so on. The study also pointed out that companionship robots with more social functions are more likely to elicit expressions and responses from the elderly and to promote the participation of the elderly in social interactions and enhance the quality of their interactions [19,50].

Limitations and Future Work

There were some limitations to this study. First, the research tool was self-reported, and the subjects' understanding of the questions in the questionnaire and their personal perception of their own situation all affected their answers. The subjects' answers might have been affected by extrinsic factors beyond their control such as mood and social expectations, so the results of the inventory might have been different from their actual situation, resulting in measurement errors in the results of the study. Second, the sampling sites for the study were gathering places for the elderly, such as community care strongholds and community centers, and at activities intended to respect the elderly and as part of Mother's Day events. The subjects were middle-aged and older adults who were more active and had the intention and ability to go out, so the results of the study are limited in terms of extrapolation. Third, the sample size and representativeness of the subjects were not as good as probability sampling, which affected any inferences that could be made based on the results, although the study areas covered the northern, central, and southern parts of Taiwan to increase the robustness of the results of the study. Fourth, this research is the first attempt at understanding companionship preferences toward pets and robots among the elderly in the community in Taiwan. Therefore, cultural factors were not considered. Similarly, because there have been few studies of this topic abroad [29], it is difficult to know whether there are similarities or differences between cultures, which is an expectant direction for future work.

In spite of these limitations, this study is a rare survey of the perceptions of elderly individuals living in the community toward companionship robots. The subjects in this study were aged 45 to 94 years and included active middle-aged and older adults living in the community in the northern, central, and southern parts of Taiwan. It is the only study in Taiwan to compare pets and robots from the point of view of choosing them as a companion object. From the results of the study, we can preliminarily understand the current situation and preferences of the needs for companionship in middle-aged and older adults living in the community. In addition to contributing to the literature, middle-aged and older adults at home can find a suitable companion object in later life based on their sociodemographic characteristics. First-line community practitioners can also design care projects for middle-aged and older adults living with different needs and backgrounds and provide a future implementation plan for the welfare system, as well as provide empirical evidence for policy promotion and the development of science- and technology-related industrial products.

Conclusion

The key findings of this study are as follows. First, variables such as age, education level, type of dwelling, occupation, retirement status, number of comorbidities, experience with keeping pets, experience with using apps, and eHealth literacy significantly affected the degree of acceptance of a robot or pet as a companion object in later life. Community practitioners working with middle-aged and older adults could plan curricula according to the different backgrounds and characteristics of the population of interest to develop care projects for middle-aged and older adults based on their different needs and backgrounds and help them to select appropriate companion objects in later life. Second, the study found that eHealth literacy significantly affected the degree of acceptance of robots and pets in the middle-aged and older adults as well as the type of functions desired in a companion robot. Those with higher eHealth literacy scores were more likely to respond that both robots and pets were acceptable as companion objects. This indicates that those with better eHealth literacy are more likely to choose a robot or pet as their companion object in later life and that eHealth literacy is significantly associated with age. First-line staff or policy makers in relevant fields can conduct eHealth literacy promotion courses for middle-aged and older adults in order to facilitate the implementation of relevant plans. Third, in terms of the development of companionship robot products, middle-aged and older women generally preferred animal-like robots as companion objects in later life, while men preferred an adult human-like robot. In terms of functions, middle-aged and older adults in the community are more likely to need a companionship robot that has functions including dancing, singing, storytelling, or news reporting. Whether or not they live alone or with a partner also affects their preferred robot functions, so technology-related industries should consider designing products to suit the needs of different target groups. It would be useful to increase the number of study samples or select specific groups to carry out intensive studies. Heterogeneous populations in residential institutions or extension of the sampling sites could make the results more generalizable. In addition, qualitative study methods such as in-depth interviews would help lead to an understanding of the needs of middle-aged and older adults in terms of a companion in later life. Study methods attaching equal importance to quality and quantity in the future could potentially better reflect the current needs of middle-aged and older adults.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Different types of robots.
[DOCX File , 467 KB - jmir_v23i6e23471_app1.docx]

Multimedia Appendix 2

Questionnaires.
[DOCX File , 50 KB - jmir_v23i6e23471_app2.docx]

References

1. Khosravi P, Ghapanchi AH. Investigating the effectiveness of technologies applied to assist seniors: A systematic literature review. Int J Med Inform 2016 Jan;85(1):17-26. [doi: 10.1016/j.ijmedinf.2015.05.014] [Medline: 26216463]
2. Chiu C, Liu C. Understanding Older Adult's Technology Adoption and Withdrawal for Elderly Care and Education: Mixed Method Analysis from National Survey. J Med Internet Res 2017 Nov 03;19(11):e374 [FREE Full text] [doi: 10.2196/jmir.7401] [Medline: 29101093]
3. Chiu C, Hu Y, Lin D, Chang F, Chang C, Lai C. The attitudes, impact, and learning needs of older adults using apps on touchscreen mobile devices: Results from a pilot study. Comput Hum Behav 2016 Oct;63:189-197. [doi: 10.1016/j.chb.2016.05.020]
4. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
5. Law M, Sutherland C, Ahn HS, MacDonald BA, Peri K, Johanson DL, et al. Developing assistive robots for people with mild cognitive impairment and mild dementia: a qualitative study with older adults and experts in aged care. BMJ Open 2019 Sep 24;9(9):e031937 [FREE Full text] [doi: 10.1136/bmjopen-2019-031937] [Medline: 31551392]
6. Dünnebeil S, Sunyaev A, Blohm I, Leimeister JM, Krcmar H. Determinants of physicians' technology acceptance for e-health in ambulatory care. Int J Med Inform 2012 Nov;81(11):746-760. [doi: 10.1016/j.ijmedinf.2012.02.002] [Medline: 22397989]
7. Yeh MI. Robots and application in healthcare. Leadership Nursing 2016;17(4):3-12. [doi: 10.29494/LN.201612_17(4).0001]
8. Robinson H, MacDonald B, Broadbent E. The Role of Healthcare Robots for Older People at Home: A Review. Int J of Soc Robotics 2014 Jul 3;6(4):575-591. [doi: 10.1007/s12369-014-0242-2]
9. Broadbent E, Tamagawa R, Patience A, Knock B, Kerse N, Day K, et al. Attitudes towards health-care robots in a retirement village. Australas J Ageing 2012 Jun;31(2):115-120. [doi: 10.1111/j.1741-6612.2011.00551.x] [Medline: 22676171]
10. Broekens J, Heerink M, Rosendal H. Assistive social robots in elderly care: a review. Gerontechnology 2009;8(2):94-103. [doi: 10.4017/gt.2009.08.02.002.00]
11. Arras K, Cerqui D. Do we want to share our lives and bodies with robots? A 2000 people survey. Tech Rep 2005:13-19.
12. Ejdys J, Halicka K. Sustainable Adaptation of New Technology—The Case of Humanoids Used for the Care of Older Adults. Sustainability 2018 Oct 18;10(10):3770. [doi: 10.3390/su10103770]
13. Banks MR, Willoughby LM, Banks WA. Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. J Am Med Dir Assoc 2008 Mar;9(3):173-177. [doi: 10.1016/j.jamda.2007.11.007] [Medline: 18294600]
14. Sheba JK, Phuc LT, Salman AA, Kumar S, Elara MR, Martínez-García E. Development of Rehabilitative Multimodal Interactive Pet Robot for Elderly Residents. Procedia Comput Sci 2018;133:401-408. [doi: 10.1016/j.procs.2018.07.049]
15. Yu R, Hui E, Lee J, Poon D, Ng A, Sit K, et al. Use of a Therapeutic, Socially Assistive Pet Robot (PARO) in Improving Mood and Stimulating Social Interaction and Communication for People With Dementia: Study Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2015 May 01;4(2):e45 [FREE Full text] [doi: 10.2196/resprot.4189] [Medline: 25934173]
16. Wada K, Shibata T, Saito T, Sakamoto K, Tanie K. Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. 2005 Presented at: Proceedings of the 2005 IEEE International Conference on Robotics and Automation; 2005; Barcelona, Spain. [doi: 10.1109/robot.2005.1570535]
17. Marti P, Bacigalupo M, Giusti L, Mennecozzi C, Shibata T. Socially assistive robotics in the treatment of behavioural and psychological symptoms of dementia. 2006 Presented at: The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006. BioRob 2006; 2006; Pisa, Italy. [doi: 10.1109/biorob.2006.1639135]
18. Huschilt J, Clune L. The use of socially assistive robots for dementia care. J Gerontol Nurs 2012 Oct;38(10):15-19. [doi: 10.3928/00989134-20120911-02] [Medline: 22998095]
19. Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP. Scoping review on the use of socially assistive robot technology in elderly care. BMJ Open 2018 Feb 12;8(2):e018815 [FREE Full text] [doi: 10.1136/bmjopen-2017-018815] [Medline: 29440212]
20. Coşar S, Fernandez-Carmona M, Agrigoroaie R, Pages J, Ferland F, Zhao F, et al. ENRICHME: Perception and Interaction of an Assistive Robot for the Elderly at Home. Int J Soc Robotics 2020 Feb 01;12(3):779-805. [doi: 10.1007/s12369-019-00614-y]
21. Shibata T, Wada K. Robot therapy: a new approach for mental healthcare of the elderly - a mini-review. Gerontology 2011;57(4):378-386 [FREE Full text] [doi: 10.1159/000319015] [Medline: 20639620]
22. Stern C. The meaningfulness of Canine-Assisted Interventions (CAIs) on the health and social care of older people residing in long term care: a systematic review. JBI Libr Syst Rev 2011;9(21):727-790. [doi: 10.11124/01938924-201109210-00001] [Medline: 27820415]
23. Garrity TF, Stallones LF, Marx MB, Johnson TP. Pet Ownership and Attachment as Supportive Factors in the Health of the Elderly. Anthrozoös 2015 Apr 27;3(1):35-44. [doi: 10.2752/089279390787057829]

24. Chowdhury E, Nelson M, Jennings G, Wing L, Reid C. Pet ownership and survival in the elderly hypertensive population. J Hypertens 2017;35(4):769-775. [doi: 10.1097/hjh.0000000000001214]
25. Filan SL, Llewellyn-Jones RH. Animal-assisted therapy for dementia: a review of the literature. Int Psychogeriatr 2006 Apr 26;18(4):597-611. [doi: 10.1017/s1041610206003322]
26. Mordoch E, Osterreicher A, Guse L, Roger K, Thompson G. Use of social commitment robots in the care of elderly people with dementia: a literature review. Maturitas 2013 Jan;74(1):14-20. [doi: 10.1016/j.maturitas.2012.10.015] [Medline: 23177981]
27. Yeh M, Liao H, Chen S. Animal-Assisted Therapy in Taiwan. J Formos Med Assoc 2002;6(1):102-106.
28. 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]
29. Thodberg K, Sørensen LU, Videbech PB, Poulsen PH, Houbak B, Damgaard V, et al. Behavioral Responses of Nursing Home Residents to Visits From a Person with a Dog, a Robot Seal or a Toy Cat. Anthrozoös 2016 Mar 08;29(1):107-121. [doi: 10.1080/08927936.2015.1089011]
30. Hu HF, Wang CH, Chang SM, Huang HC, Lai ZY, Sung HC. Preliminary Study on the Effects of Robot-Assisted Therapy on Depression and Agitated Behaviors among Older People with Dementia. VGH Nursing 2014;31(4):379-387. [doi: 10.6142/VGHN.31.4.379]
31. Hu HF, Sung HC. A Nursing Experience of Using a Robot-assisted Therapy for Behavioral and Psychological Symptoms of an Institutionalized Older Adult with Dementia. J Long-Term Care 2012;16(1):91-102. [doi: 10.6317/LTC.201204_16(1).0006]
32. Leng M, Liu P, Zhang P, Hu M, Zhou H, Li G, et al. Pet robot intervention for people with dementia: A systematic review and meta-analysis of randomized controlled trials. Psychiatry Res 2019 Jan;271:516-525. [doi: 10.1016/j.psychres.2018.12.032] [Medline: 30553098]
33. Cruz-Sandoval D, Morales-Tellez A, Sandoval EB, Favela J. A social robot as therapy facilitator in interventions to deal with dementia-related behavioral symptoms. 2020 Presented at: 2020 ACM/IEEE International Conference on Human-Robot Interaction; March 23-26, 2020; UK. [doi: 10.1145/3319502.3374840]
34. Health Promotion Administration. Long-term follow-up survey on the physical and mental social life of middle-aged and elderly people. Health Promotion Administration. 2011. URL: https://www.hpa.gov.tw/Pages/List.aspx?nodeid=108 [accessed 2018-01-15]
35. Barrera M, Sandler IN, Ramsay TB. Preliminary development of a scale of social support: Studies on college students. Am J Community Psychol 1981;9(4):435-447. [doi: 10.1007/bf00918174]
36. Luo KN. A Study on the Influence of Social Support, Personality Traits, and Personal Attributes on the Psychological Well-Being of the Elderly [master thesis]. Taipei, Taiwan: Department of Psychology, National Cheng Chi University; 2001.
37. Li SY. The Relationship of Resources loss, Social Support, Coping Style, and Psychological Well-Being among Middle-Aged and Older People in Taiwan [master thesis]. Taichung, Taiwan: Department of Psychology, Chung Shan Medical University; 2013:1-132.
38. Radloff LS. The CES-D Scale. Appl Psychol Meas 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
39. Lee KL, Ou YL, Chen SH. The Psychometric Properties of a Short Form of the CES-D used in the Taiwan Longitudinal Study on Aging. Formosa Journal of Mental Health 2009;22(4):383-410.
40. Goldberg LR. The development of markers for the Big-Five factor structure. Psychol Assess 1992;4(1):26-42. [doi: 10.1037/1040-3590.4.1.26]
41. Zheng L, Goldberg LR, Zheng Y, Zhao Y, Tang Y, Liu L. Reliability and Concurrent Validation of the IPIP Big-Five Factor Markers in China: Consistencies in Factor Structure between Internet-Obtained Heterosexual and Homosexual Samples. Pers Individ Dif 2008 Nov 01;45(7):649-654 [FREE Full text] [doi: 10.1016/j.paid.2008.07.009] [Medline: 20383283]
42. Li RH, Chen YC. The Development of a Shortened Version of IPIP Big Five Personality Scale and the Testing of Its Measurement Invariance between Middle-Aged and Older People. Journal of Educational Research and Development 2016;12(4):87-119. [doi: 10.3966/181665042016121204004]
43. 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]
44. Cheng SY, Chang FC, Li JM. eHealth Literacy and Related Factors among Junior High School Students in Taipei City. J Health Promot Health Educ 2014;41:1-24.
45. Arras K, Cerqui D. Do we want to share our lives and bodies with robots? A 2000 people survey. Tech Rep 2005:605-605.
46. Yu R, Hui E, Lee J, Poon D, Ng A, Sit K, et al. Use of a Therapeutic, Socially Assistive Pet Robot (PARO) in Improving Mood and Stimulating Social Interaction and Communication for People With Dementia: Study Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2015 May 01;4(2):e45 [FREE Full text] [doi: 10.2196/resprot.4189] [Medline: 25934173]
47. Huang YC, Deng YS. Design Guidelines of Companion Robot for Elderly People [thesis]. Hsinchu, Taiwan: National Chiao Tung University; 2007.

48. DiSalvo C, Gemperle F, Forlizzi J, Montgomery E. The hug: an exploration of robotic form for intimate communication. 2003 Presented at: The 12th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN); 2003; Millbrae, CA, USA p. 403-408. [doi: 10.1109/roman.2003.1251879]
49. Forlizzi J, DiSalvo C, Zimmerman J, Mutlu B, Hurst A. The SenseChair: The lounge chair as an intelligent assistive device for elders. 2005 Presented at: The Conference on Designing for User eXperience; 2005; San Francisco, California, USA.
50. Heerink M, et al. Enjoyment intention to use and actual use of a conversational robot by elderly people. In: Proceedings of the 3rd ACM/IEEE international conference on human robot interaction. 2008 Presented at: 3rd ACM/IEEE international conference on human robot interaction; March 2008; Amsterdam, The Netherlands.

Abbreviations

CES-D: Center for Epidemiologic Studies Depression Scale
eHEALS: eHealth Literacy Scale
ICT: information and communications technology
IPIP: International Personality Item Pool
IPIP-15: 15-item version of the International Personality Item Pool
IRB: institutional review board
ISSB: Inventory of Socially Supportive Behavior
mHealth: mobile health

Edited by G Eysenbach; submitted 13.08.20; peer-reviewed by P Anderberg, F Saigí-Rubió, R Zhong; comments to author 02.12.20; revised version received 24.03.21; accepted 13.04.21; published 11.06.21.

Please cite as:
Chiu CJ, Hsieh S, Li CW
Needs and Preferences of Middle-Aged and Older Adults in Taiwan for Companion Robots and Pets: Survey Study
J Med Internet Res 2021;23(6):e23471
URL: https://www.jmir.org/2021/6/e23471/
doi:10.2196/23471
PMID:34347621

Viewpoint

Patient-Centered Care: Transforming the Health Care System in Vietnam With Support of Digital Health Technology

Thu Ha Dang[1,2,3], MD, MIPH; Tuan Anh Nguyen[1,4,5,6], BPharm, MPharm, PhD; Minh Hoang Van[7,8], MD, MPH, PhD; Olinda Santin[9], PhD; Oanh Mai Thi Tran[6], MD, PhD; Penelope Schofield[2,10,11], BSc (Hons), PhD

[1]Department of Psychological Sciences, School of Health Sciences, Swinburne University of Technology, Melbourne, Australia
[2]Behavioural Sciences Unit, Health Services Research and Implementation Sciences, Peter MacCallum Cancer Centre, Melbourne, Australia
[3]Digital Health Cooperative Research Centre, Sydney, Australia
[4]Social Gerontology Division, National Ageing Research Institute, Melbourne, Australia
[5]Quality Use of Medicines & Pharmacy Research Centre, Clinical and Health Sciences, University of South Australia, Adelaide, Australia
[6]Health Strategy and Policy Institute, Hanoi, Vietnam
[7]Hanoi University of Public Health, Hanoi, Vietnam
[8]National Institute of Health Sciences, Bach Mai Hospital, Hanoi, Vietnam
[9]School of Nursing and Midwifery, Medical Biology Centre, Queen's University Belfast, Belfast, United Kingdom
[10]Department of Psychology, and Iverson Health Innovation Research Institute, Swinburne University of Technology, Melbourne, Australia
[11]Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Australia

Corresponding Author:
Thu Ha Dang, MD, MIPH
Department of Psychological Sciences
School of Health Sciences
Swinburne University of Technology
John Street
Hawthorn
Melbourne, 3122
Australia
Phone: 61 422703347
Email: thuhadang@swin.edu.au

Abstract

Background: Over the recent decades, Vietnam has attained remarkable achievements in all areas of health care. However, shortcomings including health disparities persist particularly with a rapidly aging population. This has resulted in a shift in the disease burden from communicable to noncommunicable diseases such as dementia, cancer, and diabetes. These medical conditions require long-term care, which causes an accelerating crisis for the health sector and society. The current health care system in Vietnam is unlikely to cope with these challenges.

Objective: The aim of this paper was to explore the opportunities, challenges, and necessary conditions for Vietnam in transforming toward a patient-centered care model to produce better health for people and reduce health care costs.

Methods: We examine the applicability of a personalized and integrated Bespoke Health Care System (BHS) for Vietnam using a strength, weakness, opportunity, and threat analysis and examining the successes or failures of digital health care innovations in Vietnam. We then make suggestions for successful adoption of the BHS model in Vietnam.

Results: The BHS model of patient-centered care empowers patients to become active participants in their own health care. Vietnam's current policy, social, technological, and economic environment favors the transition of its health care system toward the BHS model. Nevertheless, the country is in an early stage of health care digitalization. The legal and regulatory system to protect patient privacy and information security is still lacking. The readiness to implement electronic medical records, a core element of the BHS, varies across health providers and clinical practices. The scarcity of empirical evidence and evaluation regarding the effectiveness and sustainability of digital health initiatives is an obstacle to the Vietnamese government in policymaking, development, and implementation of health care digitalization.

Conclusions: Implementing a personalized and integrated health care system may help Vietnam to address health care needs, reduce pressure on the health care system and society, improve health care delivery, and promote health equity. However, in order to adopt the patient-centered care system and digitalized health care, a whole-system approach in transformation and operation with a co-design in the whole span of a digital health initiative developing process are necessary.

(J Med Internet Res 2021;23(6):e24601) doi:10.2196/24601

KEYWORDS

building blocks; digital health; eHealth; patient-centered care; telemedicine; Vietnam

Introduction

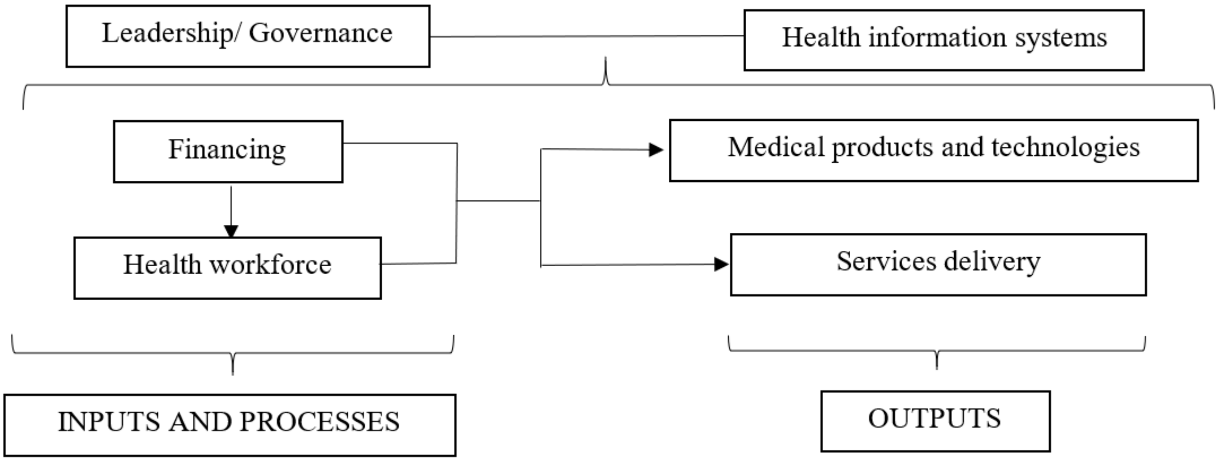
Overview

Following broad economic reforms known as Doi Moi in 1986, Vietnam has attained remarkable health care improvement, reflected in core health indicators [1]. From 1988 to 2018, life expectancy at birth increased from 69.9 years to 75.3 years, under-5 mortality rate decreased from 56‰ to 20.7‰, and infant mortality reduced from 39.6‰ to 16.5‰ [2]. Health care expenditure gradually increased and was forecasted to triple from US $15.6 billion in 2018 to US $42.9 billion in 2028 [3]. Despite these improvements, the health care system still faces significant challenges including wide disparities in health and growing health care costs. The disparities in core health indicators are particularly observed between urban and rural residents, across different regions, and among population groups [4]. For example, the maternal mortality ratio and infant mortality rate in some mountainous areas are 3 to 4 times higher than those in lowland and urban areas and almost double the national average rates [4].

Vietnam is undergoing a dramatic demographic transition resulting in an aging population. The number of people aged 65 years and over is estimated to increase from 10% of the population in 2015 to 28% in 2050 [5]. The combination of an aging population, increased industrialization, and changes in population lifestyle have created a double disease burden, with a shift from communicable to noncommunicable diseases (NCDs). Specifically, the mortality rate caused by NCDs rose from 45.5% in 2010 to 77% in 2016 and is projected to climb [3,4,6]. The double disease burden means that Vietnam is facing more costly health conditions such as dementia, cancer, and multimorbidity. In addition to this, the country still faces significant burden of infectious conditions and a number of new epidemics such as COVID-19. These health challenges must be addressed using a systemic approach by the whole government of Vietnam to improve health for its population.

To improve health care problems and the health status of the population, there is a critical need for a well-functioning health care system that can deliver services equitably and efficiently [7]. The World Health Organization (WHO) developed an evidence-based building blocks framework as a tool to help its member states analyze their health care systems. This framework allows nations to consider the multifaceted nature of their health systems and interdisciplinary and multilevel responsibility in health care [7]. The WHO framework assesses health systems using 6 core components or building blocks. Each building block and its indicators were initiated by a group of agency representatives and technical experts, shared broadly with country experts, and followed by evaluations through a series of case studies and reviews of country experiences [7]. These components focus on the key chains of the monitoring and evaluation framework developed by the International Health Partnership, namely inputs, processes, and outputs [8]. The relation between the 6 building blocks and this monitoring and evaluation framework is summarized in Figure 1. In the following section, the 6 building blocks and corresponding indicators are used to describe Vietnam's health care system, comprising financing and health workforce components (inputs and processes), medical products, technologies, and service delivery components (immediate outputs), and cross-cutting components: leadership/governance and health information systems [7].

Figure 1. Health system strengthening: relationship between the World Health Organization building blocks (6 core components, top) and International Health Partnership monitoring and evaluation framework (inputs, processes and outputs, bottom).



Health Care System in Vietnam

Health Financing

Health financing is a key building block in a national health system, largely influencing the inputs, thus affecting the availability, affordability, and accessibility of health services. A good health financing system should move toward universal health coverage, where all people have access to needed health services without financial hardship. This could be achieved through increasing total health expenditure (THE) and decreasing the proportion of households facing financial catastrophe as a result of out-of-pocket payments (OPP) [7].

With multiple health financing reforms, Vietnam's THE per capita increased from US $14 in 1995 to US $113 in 2014 [9,10], thus within the internationally defined range and enough for universal coverage of key health interventions [7]. The increase in public health expenditure (mainly comprising state budget [11] and social health insurance [12]) has increased health care coverage for some groups including the poor, ethnic minorities, under–6-year-old children, over–80-year-old people, and socially vulnerable groups through the government's subsidized schemes [13]. However, patient OPP remained high, accounting for 40.8% of THE in 2015, which was higher than that of other countries in the Asia Pacific region and the WHO recommended level [10,11]. The high OPP led to catastrophic expenditure and pushed many Vietnamese families into poverty, resulting in health care inequity [14,15]. The current model of the health care system and financing needs further reforms to address a surge in health care expenditure caused by the aging population and shifting disease pattern in Vietnam.

Health Workforce

The health workforce is another key building block to provide inputs and processes to the monitoring and evaluation chain of health systems. The ability of a country to meet its health goals largely depends on people in charge of organizing and delivering health services. Evidence of the direct and positive link between numbers of health workers and population health outcomes has been demonstrated in several studies [16,17].

The health workforce in Vietnam has gradually improved in both quantity and quality. The number of doctors and pharmacists increased from 7.2 and 1.76 per 10,000 people in 2010 to 8.0 and 2.2 per 10,000 people in 2015, respectively [4]. In 2015, 65% to 95% of the health facilities and about 90% of the health workers in hospitals at central and provincial levels were licensed [4]. Despite this significant improvement, Vietnam's health workforce was still insufficient to meet staffing norms and clinical needs [4] and inappropriately distributed across regions and levels/areas of care. The aging population and shifting disease burden to those requiring long-term care for chronic NCDs are likely to lead to severe shortages in health resources, which occur in highly specialized fields such as cancer, palliative care, and mental health and in hard-to-reach areas such as North West, Central highlands, and Mekong Delta regions [4,18-20]. The mountainous and remote areas lack not only specialists trained in advanced diagnostic and treatment approaches but also standard medical and diagnostic equipment, which diminishes the quality of health care services in these areas compared to urban regions [4]. This wide disparity in health care between the rich and the poor, urban and rural, is demonstrated in the disparity of core health indicators such as life expectancy at birth and infant and under-5 mortality rates [4].

Overcrowding in health facilities, especially in urban and specialized hospitals, is a main cause of health worker exhaustion. Nearly one-fifth of Vietnamese clinical nurses experienced burnout and occupational stress [21]. Clinician burnout directly reduces the quality of life of clinicians and adversely affects the quality of care to patients. It also indirectly contributes to the reduction of health staffing [22]. Hence, a vicious negative cycle for Vietnam's health sector is created.

Health Service Delivery

Health service delivery is reflected in the availability and readiness of services across the health care continuum. Vietnam has achieved significant service improvement in health care: for example, the ratio of hospital beds per 10,000 people increased from 21.5 in 2011 to 24.0 in 2015 [4]. The hospital quality management system was established in 2013 and available in 55.4% of hospitals throughout the country in 2015 [23].

Service provision is immediate outputs of the inputs into the health system such as financing, workforce, procurement, and supplies [7]. It will be difficult to achieve the outputs if the inputs are insufficient. Even if the inputs are adequate, whether the outputs are obtained depends very much on the efficiency of the health system's functioning. According to Bentley et al [24], there are 4 key inefficiencies: duplication of services, inefficient processes, overly expensive inputs, and medical errors. All of these forms of inefficiency occur in Vietnam's health sector.

Vietnamese patients' laboratory tests and results are not usually archived at medical facilities or shared between different health care providers. This poses a challenge as patients experience multiple visits to doctors and specialists for the same health conditions thus leading to duplication of services [3]. Public hospitals, especially in large cities, are usually overcrowded with 2 to 3 patients sharing a bed. Limited quality of health care services at the commune level leads to reduced patient trust in primary care. Also, if patients use OPP, the health care system allows them to easily bypass lower level facilities (eg, commune health stations) and seek health care services in leading tertiary hospitals in big cities without referrals (ie, inefficient processes), even just to treat common diseases that primary care is well equipped to manage (ie, overly expensive inputs) [4,25]. Consequently, higher level hospitals are drained of resources, while there is waste at lower levels due to underuse [26]. Overcrowding in high-level health facilities is associated with medication errors. A large prospective study in two urban hospitals in Vietnam revealed that medication errors occurred in more than one-third of all medication doses [27].

There are also discrepancies in health service readiness and quality across areas of health care. Vietnam has an extensive primary health care system that reaches almost every administrative jurisdiction and acts as the main entry point to public health care. However, grassroots level facilities have inadequate infrastructure required for basic health care delivery. For example, only 76% of commune health stations in Dien Bien province (a Northern mountainous province) have a source of clean water, and a significant number of district hospitals lack essential equipment such as child ventilators and electrocardiograms [28]. Also, there have been increasing concerns about the equity and quality of basic health service provisions in primary health care. These limitations were reflected in key health indicators, with the infant mortality rate among the ethnic minority population being over 4 times higher than that of Kinh and Hoa ethnic groups, who mostly live in urban areas [28].

Access to Essential Medicines

According to WHO, a well-functioning health system should be able to provide the population it serves with equitable and affordable access to essential medicines, medical products, and technologies and use this resource efficiently [7]. However, published literature consistently reported that medicine prices in Vietnam were high and unaffordable for many Vietnamese people [29-31]. A fragmented medical information system within and between health care and other sectors such as General Department of Vietnam Customs and Ministry of Finance, as well as a shortage of personnel and resources for enforcing medicine pricing policies was one of the reasons for their high price in Vietnam [32]. Because of high medicine prices, irrational selection and use of medicines, unsustainable pharmaceutical production and distribution systems, and a lack of financial support systems for medicine procurement, access to the right medicines at the times people need them remains a major challenge for the majority of the Vietnamese people [33-37].

Health Information

The health information system (HIS) is a crosscutting building block because it creates a foundation for all decision-making processes in a health system. Vietnam has achieved significant progress in this area. A number of health statistics are generated annually: for example, the Annual Health Statistic Yearbook, Joint Annual Health Review, and Statistical Yearbook. Information technology is applied widely in the health administration and management of all 63 provinces and cities across the country. Vietnam is also strongly promoting the development of a health management database for its over 90 million citizens [4].

Vietnam's HIS, however, still faces a number of issues and challenges, including data generation, validation, and uses. For example, data from health care facilities, especially from private and industrial sectors, are neither timely nor complete. Information from local death registration is incompatible with WHO recommendations. There are discrepancies in the quality of medical records across regions and levels of care, which creates challenges for continuing health monitoring, disease prevention and treatment, and management of medical errors and adverse drug reactions. Even after the data are collected, the unclear data dissemination mechanism in Vietnam will likely restrict data use [4]. Given the importance of this crosscutting building block of the health system, reform is needed to operate HIS more effectively, especially in the new era of information technology.

Leadership and Governance

Health leadership and governance is another crosscutting building block in a health care system. It mediates other building blocks by connecting all issues surrounding the accountability of various stakeholders in the system to ensure adequate resources (finance, workforce, medical supplies, and information) are available to deliver essential health services. In terms of health care governance, Vietnam has attained remarkable progress, demonstrated in the development of a strong policy framework in health care. The enactment of important laws and policies, such as health insurance and pharmacy laws [38,39], and the Health Sector Strategy for 2011-2020 with a Vision to 2030 [40] provided a solid foundation and led to the formation of regulations, guidelines, initiatives, and plans in these areas. The organizational structure of Vietnam's health system has been adjusting to meet health care needs at various levels, such as establishing information technology administration at central levels and formally affirming the function and tasks of community health services at local levels [4]. Nevertheless, health policies in Vietnam are sometimes overlap, inconsistent and lack evidence. Lack of detailed plans and information cause difficulty for the effective implementation of policies. Additionally, insufficient sanctions and a weak inspection network for policy enforcement are also shortcomings of Vietnam's health care governance [4].

In summary, Vietnam has attained significant achievement and improvement in all 6 areas of the health system framework. However, shortcomings persist. To address all these systemic challenges and achieve optimal and equitable health outcomes for the population, the Vietnamese government must consider adapting the current model of health care operation to meet rapidly changing population trends, patterns of disease, and health care needs and use existing resources more efficiently while improving health infrastructure. The application of digital health technologies to underpin the health care system transformation is critical to success [41]. "The use and scale-up of digital health solutions can revolutionize how people worldwide achieve higher standards of health and access services to promote and protect their health and well-being" [42]. As in many other countries, Vietnam's health care system will need to be migrated to the patient-centered care model, which focuses on patients and their particular health care needs, and patients will need to be empowered to become active participants in their care to optimize health and economic outcomes [43]. This study was conducted to explore the opportunities, challenges, and necessary conditions for Vietnam in transforming toward a patient-centered care model to produce better health for people and reduce health care costs.

Methods

Approach

In this paper, we explored a Bespoke Health Care System (BHS) developed by Schofield et al [44] as an ideal example of a comprehensive patient-centric care model. The current Vietnamese digital health and health care landscape was examined using a strength, weakness, opportunity, and threat (SWOT) analysis to argue for the potential of developing a BHS in Vietnam. We then discuss necessary and sufficient conditions and challenges for Vietnam in transforming toward this patient-centered care model to produce better health for people and reduce health care costs.

Bespoke Health Care System

In 2019, Schofield et al [44] first introduced the concept of the BHS in the context of Australian health care, which has been adapted to other contexts [45,46]. The development of the BHS was based on the pedagogical model of flipping the classroom in modern education [47] and its application in health care [48]. In Australia, this approach has been applied by many health professionals [47]. Patients were equipped with necessary knowledge before consultations. Therefore, the consultation time was effectively used to solve health problems and make joint decisions [47]. The core component of the BHS is "increasing patient involvement in health care decisions and self-management assisted by the use of technology" [49]. In this model, patients will be educated about their illness and management options, flipping their role from passengers to drivers to manage their own health care, with clinicians playing more of a support role being the "guide by the side." Although the BHS was proposed as an ideal comprehensive patient-centric care system with several advantages and benefits, there is no one-size-fits-all approach. Novel health care solutions will work best when they are adapted to suit a country's specific conditions and have broad-based acceptance among the community, health care providers, and government agencies.

SWOT Analysis

The SWOT analysis has been used widely in policy research to provide policy makers with a sound basis for strategy development and formulation and identify new avenues for national health care reform [50]. The technique examines 4 parameters: strengths, weaknesses, opportunities, and threats. In this paper, strengths and weaknesses refer to internal factors of the proposed BHS that place it at an advantage or disadvantage over the current Vietnamese health care system, respectively. Opportunities and threats refer to external factors to the BHS that support or prevent it from being adopted in Vietnam.

Results

Strengths

The BHS brings a number of potential benefits to a health care system. First, the BHS proposes integrating electronic medical records (EMRs) into a patient-centered management platform. Currently, EMRs act as passive information depositories mainly for the purposes of health data storage or analysis. In the proposed BHS, every person would have their EMR containing all relevant personal and medical information that is shared across health care providers. As such, relevant health workers and agencies would have access to patients' real-time health information, saving time in managing people's health, saving costs of unnecessary or duplicated examinations and laboratory tests, and reducing medication errors. Studies have revealed that one-fifth of medical errors were due to insufficient patient medication information [48,51]. Clinical access to patients' real-time health information would enable tailored precision and holistic health solutions for individuals.

Second, the BHS proposes that the EMR is an active tool in promoting optimal health management. The proposed platform would remotely track symptoms, prompt patients to perform routine screening tests, allow patients to book medical appointments and remind them to attend, and provide scheduled reminders for adherence to medical recommendations. Electronic prompts and reminder alerts have been shown to assist individuals in adhering to clinical intervention effectively, particularly in managing long-term chronic conditions [49].

Third, the BHS serves as an education platform, upskilling patients and health workers. In the BHS, patients would optimally be equipped with a comprehensive understanding of their conditions, treatment options, and self-management strategies. This is an effective way to empower people to have greater ownership in managing their health. Additionally, clinicians would be provided with up-to-date, evidence-based optimal health care pathways suitable for their specific patients' conditions. This offers an ongoing mechanism to support clinicians' continuing professional development using their own case studies. Moreover, the platform might be used to upskill patient families and carers, which would strengthen community care and lessen the health care burden due to staff shortages [51]. Potentially, the BHS would facilitate reducing the health inequities in terms of accessing optimal health care, particularly for the economically disadvantaged or those living in remote areas.

In summary, adopting the BHS model may assist Vietnam in addressing the health crisis and achieving the country's health care goals in the new decade. This model can increase patient accessibility to health care facilities and state-of-the-art health management, including the most vulnerable and hard-to-reach people, thus enhancing health equity across the country. It can also increase operational efficiencies for both health care providers and users, resulting in lessening overcrowded hospitals and enabling coordinated health care, which is currently missing [3,52]. Moreover, digital solutions can assist teaching, tertiary, or specialized hospitals to deliver training or conduct eHealth consultations with satellite, primary, or secondary hospitals. Therefore, the quality of health care would be strengthened across the country.

Weaknesses

Although the BHS model has many strengths, we need to articulate inherent weaknesses in the model. First, the high cost and complexity of implementing digital health information systems, such as EMRs, may be a barrier to broad dissemination. Second, the digitalization of data and services represents a potential cybersecurity threat to privacy and trust of people in a new health care system [3]. Finally, the establishment of nationwide unique health IDs (unique codes used to identify individuals within the health care system) often takes time, especially for socially disadvantaged members of the community.

Opportunities

Vietnam has a high-level policy framework (ie, political and legal environment) supporting the transition toward the BHS. The 2017 Resolution 20-NQ/TW of the Communist Party of Vietnam has provided a strategic orientation for reforming the health care system. This includes systemic implementation of information technology in management of primary health care, prevention, disease management, and the establishment of electronic health records (EHRs) for all citizens that link to their health insurance card.

Based on this strategic orientation, the Vietnamese Ministry of Health (MOH) in 2017 set out national goals for the protection, care, and improvement of people's health in the period to 2030. One of the 5 key priorities for action to achieve the national health care goals was developing human resources, medical sciences, and technology [4,53]. Since then, a national agenda for digitalization of the health care system has been driven by MOH with a number of initiatives that aim to adopt digital solutions in health care network across the country. Vietnam is rapidly embracing digital health. Digitalized health care is perceived as a useful solution that could help to address the rapidly growing gap between service demand and capacity [44].

To implement health care digitalization, MOH issued Circular 54/2017/TT-BYT regarding assessment criteria for information technology applications in health care facilities [50] and a plan to develop smart health care during the period 2018-2025, with a vision toward 2030 [54]. In this plan, targeted smart health outcomes were specified, such as development of a national health care data center, electronic health and medical records, and electronic government and smart medicine management systems. Particularly in the field of examination and treatment, telemedicine is regulated by Circular 49/2017/TT-BYT, which provides guidelines for a range of telemedicine consultations in Vietnam (eg, allowing doctors to provide telemedicine services to patients under certain infrastructure requirements [55]).

To accelerate the application of information technology in the health sector, in 2018, MOH issued Circular 46/2018/TT-BYT regulating EMRs. According to Häyrinen et al [56], EMR was defined as "a repository of patient data in digital form, stored and exchanged securely, and accessible by multiple authorized users. It contains retrospective, concurrent, and prospective information, and its primary purpose is to support continuing, efficient, and quality integrated health care." In the circular, EMRs include inpatient records, outpatient records, and other medical records prescribed by MOH, such as traditional medicine medical records and medical records of abortion. This is the first time Vietnam has specific guidance for EMR establishment, use, and management at health care facilities [57]. As the health care system in Vietnam is centralized under MOH and provincial health authorities, the integration of EMRs within the existing system, at least in the 47 central-level hospitals and possibly also at 419 provincial-level hospitals [58], will be easier than in countries with a more fragmented health care system.

The plan for implementing EHRs was approved by MOH in 2019 in Decision 5349/QĐ-BYT [59]. According to this decision, EHRs are medical documents that record the health care process of a person from birth to death in the form prescribed by MOH. This plan guides and directs all 61 cities and provinces of Vietnam to simultaneously deploy and make EHRs available for use by 80% and 95% of population in 2020 and 2025, respectively. If this aim is achieved, Vietnam will be perfectly positioned to adapt and adopt the BHS model. EHRs may be able to assist every citizen to understand and manage their health information continuously for life, enabling them to prevent disease proactively, and actively manage their health conditions if used as the basis for a BHS. EHRs will also support physicians in providing health care for people in real time and according to best practice health care recommendations. Importantly, EHRs will provide complete, accurate, and timely data on population health, enabling policy makers to advance evidence-based health policy and health authorities to examine the relative efficacy of treatments to manage health care expenditure more efficiently.

Supportive social and technological environments also place Vietnam in a good position to adopt advanced digital health solutions of the BHS. Today, Vietnam is benefiting from a golden population proportion with 70% of the population being of working age (aged 18 to 65 years) [6]. This generation is rapidly embracing new communication technologies. In 2020, Vietnam was the seventh highest number of Facebook users in the world [60]. On average, Vietnamese people spend about 7 hours per day on the internet [61]. The high uptake is associated with a high acceptability of digital technologies by Vietnamese people. Studies have shown that people's attitudes toward mobile health solutions is highly positive. Two-thirds of Vietnamese youth and adolescents found mobile health apps useful [62,63].

Online communication is backed by a rapid and strong development of the country's information and communication technology infrastructure: penetration rate of internet access is 67%, which is among the highest in the Asia Pacific region [64], and 95% of households are now able to use 4G network [3]. Vietnam's technology infrastructure is also embracing cloud-based services, which will facilitate innovative and cost-effective health care delivery solutions. In recent years, a number of projects have initiated and implemented digital health services in large urban hospitals, including teleradiology, teleconsultation, telediagnosis, and videoconferencing. Examples of these initiatives include raising disease awareness and encouraging people to adopt heathier behaviors [65-69], improving accessibility to health care services among disadvantaged and vulnerable target groups [70-72], and upskilling health care workers [73-80].

A supportive economic environment has resulted in an increased need for high-quality health care services and precision medicine. Strong economic growth with an average annual gross domestic product growth rate of 6.4% [81] is creating a booming middle class in the society. It is estimated that this population will increase sharply from 10% in 2015 to more than 50% in 2035 [82]. The growth in disposable income among digitally literate, educated, and wealthy individuals creates economic conditions for higher personal expenditure on easily accessible, high-quality health care. Public health facilities often do not meet their needs due to long waiting times, inadequate consultation time, overcrowding, and high-occupancy bed rates [3]. This has resulted in the rapid expansion of a private health system, which was projected to grow from US $6.6 billion in 2016 to US $8.7 billion in 2020 [83]. Smart solutions that use big data, cloud computing, and mobile technology are strongly encouraged to alleviate overcrowded public hospitals and increase quality health care overall [4].

Together, all these supportive policy, social, technological, and economic environments provide a good foundation for the health care system in Vietnam to shift toward innovative models of care that are being proposed in developed countries, such as the BHS, which focus on patient-centered care.

Threats

In Vietnam, patient medical records remain paper-based at all levels of the health care system and currently are required for legal purposes [84]. Moreover, although electronic administration management has been adopted in some large hospitals, the quality of medical records and databases varies across hospitals and clinics. In most health care facilities, the medical record system is not centralized [52]. As such, one patient can have multiple medical records. This is a barrier for timely medical information exchange and sharing between hospitals [69]. In addition, the development of EMRs and EHRs, a precondition for BHS adoption, is still in the early stages [4,44]. Although MOH has taken initial steps toward the development of EMR and EHR systems, the readiness to implement the EMR nationally in real clinical practice still requires considerable preparatory effort.

First, Vietnam needs to assign a unique patient identifier to every citizen, which is a core requirement for successfully introducing EHRs [52]. In 2019, MOH issued Decision 4376/QĐ-BYT, regulating the establishment, use, and management of IDs [85]. According to the decision, an ID will be created, comprising two series of numbers separated by a dot. The first series is the social insurance code of the individual while the second series is a product of an algorithm of administrative information including the social insurance code, full name, date of birth, gender, and place of birth. The social insurance, however, is a compulsory income protection insurance for employed people only to fully or partially offset their income that is reduced or lost due to sickness, maternity, labor accident, occupational disease, retirement, or death [86]. Although Vietnam has made significant progress in expanding social insurance coverage in recent years, enrollment rates are still low, especially in small- and medium-size enterprises due to multiple barriers such as cost of contribution, lack of trust in the government system, and administrative factors [87]. Despite large increases in government subsidies, low enrollment rates in compulsory social and health insurance still persist [88]. In May 2020, only 27% of Vietnam's workforce had social insurance [89,90], which is far short of the government's target of 50% social insurance coverage [87] and 80% health insurance coverage by 2020 [91,92]. The IDs, when established, therefore, will only cover a small proportion of the population.

Second, although many health workers and patients are aware of the advantages of using EMRs and EHRs, there are concerns about the risks of privacy violation. The Law on Network Information Security No. 86/2015/QH13, enacted in 2015, includes a requirement of "providing an adequate level of protection for personal data, following the technical standards for protection of personal data." However, there are no clear definitions of "an adequate level of protection" and "technical standards" [93]. As such, there is a lack of regulatory rigor and sanctions in managing data processors. Strengthening the legal and regulatory system to protect patient privacy and information security is fundamental for the success of EMR development and application.

Finally, since Vietnam is still in the very early stage of health care digitalization, there is no empirical evidence about the effectiveness and sustainability of digital health initiatives. Over the past few years, there have been a growing number of organizations and health startups delivering digital health solutions to improve the quality of medical services. For example, eHospital software, developed by the Corporation for Financing and Promoting Technology, was first launched in 2000 [94]. This is a comprehensive hospital management system providing supports to manage all activities relating to patients in health care facilities. In 2018, eHospital with the application of the latest 4.0 technologies such as artificial intelligence and big data was introduced. Until now, this system has been used in more than 400 hospitals and clinics in Vietnam. [95].

Another example is the Hospital Information System of the Vietnam Posts and Telecommunications Group, which was introduced in 2015. This solution with its three levels of management—state, hospital and patient—aims at supporting provincial health authorities and hospital managers to make well-informed decisions for health care improvement and assisting patients to facilitate and adhere to their health care appointments online. This allegedly leads to increasing transparency and reducing overcrowding in health care facilities [96]. Nevertheless, these initiatives have not been rigorously evaluated. Lack of evaluation results will prevent the government from evidence-based policymaking and hinder the broader implementation of existing projects or development of new initiatives.

Discussion

Necessary Conditions for Successful Adoption of the BHS Model in Vietnam

Vietnam has identifiable opportunities to adopt the BHS model and implement digitalization in health care. To grasp these opportunities, the following strategies are recommended.

First, establishing a national ID for each individual based on their social insurance code is one core requirement for successfully introducing EHRs. In a recent Organisation for Economic Co-operation and Development policy report, recommendations have been proposed for the Vietnamese government to increase the enrollment rates of social insurance. They include reducing obstacles to participation (eg, simplifying administration processes for paying insurance), introducing incentive scheme for employees to participate (eg, providing government subsidies for participation of low-income people in the voluntary system), and adopting a systematic approach to social protection (eg, considering the interaction between various social protection mechanisms) [87].

Second, successful adoption of the BHS requires a whole-system approach involving the support of different sectors in the society. Obviously, MOH would have to play a key role in this adoption process but would need to collaborate with related ministries—in particular, the Ministry of Science and Technology and the Ministry of Information and Communications—in developing and implementing a digital health strategy. The resources needed to improve the health care system are sizeable. MOH would need to garner funding from different sources, both domestic and foreign agencies: government, social entrepreneurs, or private businesses [69,70].

Third, integrating personal data in EHR and EMR systems will increase the risk of privacy violation and cybersecurity breaches. The Vietnamese government needs to improve the security of information technology platforms in general and health care in particular to protect patient privacy and information security [70].

Fourth, to improve the adoption of evidence-based practices, it will be necessary to provide resources to demonstrate the effectiveness and impact of digital health initiatives and establish a network of collaborators including health care administrators, clinicians, community representatives, digital health researchers, information technology developers, and public health education experts [69]. This will be a great opportunity for further enhancing strong collaborations between multisectors and multistakeholders at various levels, which are essential for successfully reforming the health care system in Vietnam. BHS implementation would require the involvement of local educators to provide education for communities, raising awareness of the benefits the system could bring to individuals and society, coaching to enhance self-management behaviors, and increasing engagement with the system [97]. Strong evidence about the effectiveness of digital health initiatives in Vietnam will encourage the government to develop appropriate health policies and increase opportunities for ongoing projects to scale-up by attracting funding and support [70]. Thus, there is a substantial need for further research in this area in the future.

Finally, to increase the acceptability and feasibility of digital health initiatives, a co-design approach is crucial. Co-design is defined as the “collective creativity as it is applied across the whole span of a design process” [98]. In the design process, diverse experts such as researchers, designers or developers and potential customers and users work closely together to understand the needs and preferences of end users [99,100]. A co-design approach is likely to increase scalability and dissemination of the initiative [99].

Conclusions

The overarching goal of the BHS is, with the support of digital technologies, to deliver best practice health care and reduce pressure on the current health care system by empowering people to direct their own health care, regardless of their geographical location and economic status [44]. The BHS offers a promising and intelligent health care that which may be efficient and suitable for Vietnam in the new era of the Fourth Industrial Revolution [101]. Vietnam has tremendous opportunities and a favorable policy, social, technological, and economic environment to adopt this model of comprehensive patient-centric health care. In order to uptake, adapt, and implement the BHS successfully, Vietnam needs to apply a whole-system approach in transformation and operation processes.

Acknowledgments

We thank Dr Hilary Schofield for her proofreading and thorough editing.

Conflicts of Interest

None declared.

References

1. The World Bank in Vietnam. Hanoi: The World Bank; 2019. URL: https://www.worldbank.org/en/country/vietnam/overview [accessed 2020-05-24]
2. The World Bank indicators. Washington: World Bank; 2020. URL: https://data.worldbank.org/indicator/ [accessed 2020-08-16]
3. Digital health in Vietnam: a guide to market. Australian Trade and Investment Commission. 2019. URL: https://www.austrade.gov.au/ArticleDocuments/4569/Digital%20Health%20in%20Vietnam%20Report.pdf.aspx [accessed 2021-05-23]
4. Plan for people’s health protection, care and promotion 2016-2020. Hanoi: Ministry of Health of Vietnam; 2016. URL: http://faolex.fao.org/docs/pdf/vie179587.pdf [accessed 2021-05-23]
5. World population prospects: the 2015 revision. Department of Economic and Social Affairs, Population Division, United Nations. 2015 Jun 12. URL: https://esa.un.org/unpd/wpp/Publications/Files/Key_Findings_WPP_201pdf [accessed 2020-05-24]
6. Statistical yearbook of Vietnam 2018. Hanoi: General Statistics of Vietnam; 2019. URL: https://www.gso.gov.vn/wp-content/uploads/2019/10/Nien-giam-2018-1.pdf [accessed 2021-05-24]

7. Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies. Geneva: World Health Organization; 2010. URL: https://www.who.int/healthinfo/systems/WHO_MBHSS_2010_full_web.pdf?ua=1 [accessed 2021-05-24]
8. Boerma T, Abou‐Zahr C, Bos E, Hansen P, Addai E, Low‐Beer D. Monitoring and evaluation of health systems strengthening: an operational framework. Geneva: World Health Organization; 2009. URL: https://www.who.int/healthinfo/HSS_MandE_framework_Nov_2009.pdf [accessed 2021-05-24]
9. Hoang VM, Oh J, Tran TA, Tran TGH, Ha AD, Luu NH, et al. Patterns of health expenditures and financial protections in Vietnam 1992-2012. J Korean Med Sci 2015 Nov;30 Suppl 2:S134-S138 [FREE Full text] [doi: 10.3346/jkms.2015.30.S2.S134] [Medline: 26617446]
10. Ministry of Health of Vietnam and Department of Planning and Finance. National health account: implemented in Vietnam period 1998-2012. Hanoi: Statistical Publishing House; 2014.
11. Health financing strategy in Asia-Pacific region (2010-2015). Geneva: World Health Organization; 2009. URL: https://apps.who.int/iris/bitstream/handle/10665/207746/9789290614586_eng.pdf?sequence=1&isAllowed=y [accessed 2021-05-24]
12. Health financing in Vietnam. Geneva: World Health Organization URL: https://www.who.int/vietnam/health-topics/health-financing [accessed 2020-08-22]
13. The Prime Minister of Vietnam. Decision of the Prime Minister No. 139/2002/QD-TTG on examination and treatment of poverty people. Hanoi: Legal library; 2002. URL: https://thuvienphapluat.vn/van-ban/the-thao-y-te/Quyet-dinh-139-2002-QD-TTg-kham-chua-benh-cho-nguoi-ngheo-50086.aspx [accessed 2020-08-22]
14. Xu K, Evans DB, Kawabata K, Zeramdini R, Klavus J, Murray CJL. Household catastrophic health expenditure: a multicountry analysis. Lancet 2003 Jul 12;362(9378):111-117. [doi: 10.1016/S0140-6736(03)13861-5] [Medline: 12867110]
15. van Doorslaer E, O'Donnell O, Rannan-Eliya RP, Somanathan A, Adhikari SR, Garg CC, et al. Effect of payments for health care on poverty estimates in 11 countries in Asia: an analysis of household survey data. Lancet 2006 Oct 14;368(9544):1357-1364. [doi: 10.1016/S0140-6736(06)69560-3] [Medline: 17046468]
16. Anand S, Bärnighausen T. Health workers and vaccination coverage in developing countries: an econometric analysis. Lancet 2007 Apr 14;369(9569):1277-1285. [doi: 10.1016/S0140-6736(07)60599-6] [Medline: 17434403]
17. Speybroeck N. Reassessing the relationship between human resources for health, intervention coverage and health outcomes: the world health report 2006. Geneva: World Health Organization; 2006. URL: https://www.who.int/hrh/documents/reassessing_relationship.pdf [accessed 2021-05-24]
18. Krakauer EL, Cham NTP, Khue LN. Vietnam's palliative care initiative: successes and challenges in the first five years. J Pain Symptom Manage 2010 Jul;40(1):27-30 [FREE Full text] [doi: 10.1016/j.jpainsymman.2010.04.009] [Medline: 20619210]
19. Health at a glance Asia/Pacific 2016. Organisation for Economic Co-operation and Development. 2016. URL: https://www.oecd-ilibrary.org/health-at-a-glance-asia-pacific-2016_5jlsgjk4ktxs.pdf?itemId=%2Fcontent%2Fpublication%2Fhealth_glance_ap-2016-en&mimeType=pdf [accessed 2021-05-24]
20. Vuong DA, Van Ginneken E, Morris J, Ha ST, Busse R. Mental health in Vietnam: burden of disease and availability of services. Asian J Psychiatr 2011 Mar;4(1):65-70. [doi: 10.1016/j.ajp.2011.01.005] [Medline: 23050918]
21. Nguyen HTT, Kitaoka K, Sukigara M, Thai AL. Burnout study of clinical nurses in Vietnam: development of job burnout model based on Leiter and Maslach's Theory. Asian Nurs Res (Korean Soc Nurs Sci) 2018 Mar;12(1):42-49 [FREE Full text] [doi: 10.1016/j.anr.2018.01.003] [Medline: 29463486]
22. Shanafelt T, Dyrbye L. Oncologist burnout: causes, consequences, and responses. J Clin Oncol 2012 Apr 10;30(11):1235-1241. [doi: 10.1200/JCO.2011.39.7380] [Medline: 22412138]
23. Circular 19/2013/TT-BYT: Specifying implementation guidelines for healthcare service quality management in hospitals. Hanoi: Ministry of Health of Vietnam; 2013. URL: https://vanbanphapluat.co/circular-19-2013-tt-byt-implementation-guidelines-healthcare-service-quality-management-hospitals [accessed 2021-05-24]
24. Bentley TGK, Effros RM, Palar K, Keeler EB. Waste in the U.S. health care system: a conceptual framework. Milbank Q 2008 Dec;86(4):629-659 [FREE Full text] [doi: 10.1111/j.1468-0009.2008.00537.x] [Medline: 19120983]
25. Nguyen TKT, Cheng T. Vietnam's health care system emphasizes prevention and pursues universal coverage. Health Aff (Millwood) 2014 Nov;33(11):2057-2063. [doi: 10.1377/hlthaff.2014.1141] [Medline: 25354953]
26. Le D, Kubo T, Fujino Y, Pham T, Matsuda S. Health care system in Vietnam: current situation and challenges. Asian Pac J Dis Manag 2010;4(2):23-30. [doi: 10.7223/apjdm.4.23]
27. Nguyen H, Nguyen T, van den Heuvel ER, Haaijer-Ruskamp FM, Taxis K. Medication errors in Vietnamese hospitals: prevalence, potential outcome and associated factors. PLoS One 2015;10(9):e0138284 [FREE Full text] [doi: 10.1371/journal.pone.0138284] [Medline: 26383873]
28. Demombynes G, Hurt K. Quality and equity in basic health care services in Vietnam: findings from the Vietnam District and Commune Health Facility Survey. Washington: World Bank; 2015. URL: https://microdata.worldbank.org/index.php/catalog/2728/download/39045 [accessed 2021-05-24]
29. Falkenberg T, Nguyen TB, Larsson M, Nguyen TD, Tomson G. Pharmaceutical sector in transition—a cross sectional study in Vietnam. Southeast Asian J Trop Med Public Health 2000 Sep;31(3):590-597. [Medline: 11289028]

30. Aït-Khaled N, Auregan G, Bencharif N, Camara LM, Dagli E, Djankine K, et al. Affordability of inhaled corticosteroids as a potential barrier to treatment of asthma in some developing countries. Int J Tuberc Lung Dis 2000 Mar;4(3):268-271. [Medline: 10751075]
31. Nguyen A, Knight R, Mant A, Cao M, Auton M. Medicine prices, availability, and affordability in Vietnam. Southern Med Review 2009;2(2):2-9 [FREE Full text] [doi: 10.1016/b978-0-12-813166-4.00012-7]
32. Nguyen T. Medicine Prices and Pricing Policies in Vietnam [Thesis]. Sydney: UNSWorks; 2011.
33. Nguyen T, Vitry A, Roughead E. Pharmaceutical policy in Vietnam. In: Babar Z, editor. Pharmaceutical Policy in Countries with Developing Healthcare Systems. 1st Edition. Cham: Springer International Publishing; 2017:75-94.
34. Nguyen T, Roughead E. Pharmaceutical pricing policies in Vietnam. In: Pharmaceutical Prices in the 21st Century. 1st Edition. Cham: Springer International Publishing; 2015:321-342.
35. Nguyen T, Knight R, Mant A, Razee H, Brooks G, Dang T. Inflated medicine prices in Vietnam: a qualitative study. Health Policy Plann Jun 2017;32(5):647-656. [doi: 10.1093/heapol/czw174] [Medline: 28160482]
36. Nguyen TA, Knight R, Mant A, Razee H, Brooks G, Dang TH, et al. Corruption practices in drug prescribing in Vietnam: an analysis based on qualitative interviews. BMC Health Serv Res 2018 Jul 28;18(1):587 [FREE Full text] [doi: 10.1186/s12913-018-3384-3] [Medline: 30055601]
37. Wang H, Sun Q, Vitry A, Nguyen TA. Availability, price, and affordability of selected essential medicines for chronic diseases in 11 countries of the Asia Pacific Region: a secondary analysis. Asia Pac J Public Health 2017 May;29(4):268-277. [doi: 10.1177/1010539517700472] [Medline: 28397532]
38. The Congress of Vietnam. Heath insurance law. Hanoi: Legal library; 2014. URL: https://thuvienphapluat.vn/van-ban/Bao-hiem/Luat-Bao-hiem-y-te-sua-doi-2014-238506.aspx [accessed 2020-08-23]
39. The Congress of Vietnam. Pharmaceutical law. Hanoi: Legal library; 2016. URL: https://thuvienphapluat.vn/van-ban/The-thao-Y-te/Luat-Duoc-2016-309815.aspx [accessed 2020-08-23]
40. The Prime Minister of Vietnam. Health sector strategy for 2011-2020 with a vision to 2030. Hanoi: Legal library; 2013. URL: http://www.chinhphu.vn/portal/page/portal/chinhphu/hethongvanban?class_id=2&mode=detail&document_id=165437 [accessed 2020-09-23]
41. Kay M, Santos J, Takane M. mHealth: New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011.
42. Digital health. Geneva: World Health Organization; 2020. URL: https://www.who.int/health-topics/digital-health#tab=tab_1 [accessed 2020-08-23]
43. Reynolds A. Patient-centered care. Radiol Technol 2009;81(2):133-147. [Medline: 19901351]
44. Schofield P, Shaw T, Pascoe M. Toward comprehensive patient-centric care by integrating digital health technology with direct clinical contact in Australia. J Med Internet Res 2019 Jun 04;21(6):e12382 [FREE Full text] [doi: 10.2196/12382] [Medline: 31165713]
45. Bergmann J, Sams A. Flip Your Classrom: Reach Every Student in Every Class Every Day. Washington: International Society for Technology in Education; 2019.
46. Lage MJ, Platt GJ, Treglia M. Inverting the classroom: a gateway to creating an inclusive learning environment. J Econ Educ 2000;31(1):30. [doi: 10.2307/1183338]
47. Mate KS, Salinas G. Flipping primary health care: a personal story. Healthc (Amst) 2014 Dec;2(4):280-283. [doi: 10.1016/j.hjdsi.2014.10.003] [Medline: 26250637]
48. Roughhead L, Semple S, Rosenfeld E. Literature review: medication safety in Australia. Sydney: Australian Commission on Safety and Quality in Health Care; 2013. URL: https://www.safetyandquality.gov.au/sites/default/files/migrated/Literature-Review-Medication-Safety-in-Australia-2013.pdf [accessed 2021-05-27]
49. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
50. Circular promolgating the criteria of application of information technology application at medical examination and treatment establishments. Hanoi: Ministry of Health of Vietnam; 2017. URL: https://vanbanphapluat.co/circular-54-2017-tt-byt-assessment-of-information-technology-application-at-health-facilities [accessed 2021-05-24]
51. Schaller S, Marinova-Schmidt V, Setzer M, Kondylakis H, Griebel L, Sedlmayr M, et al. Usefulness of a tailored ehealth service for informal caregivers and professionals in the dementia treatment and care setting: the ehealth monitor dementia portal. JMIR Res Protoc 2016 Apr 05;5(2):e47 [FREE Full text] [doi: 10.2196/resprot.4354] [Medline: 27050401]
52. Hochwarter S, Do D, Nguyen T, Larsson M. Towards an electronic health record system in Vietnam: a core readiness assessment. J Health Inform Devel Countr 2014;8(2):1 [FREE Full text]
53. Vietnamese Party's Central Committee. Resolution 20-NQ/TW. Hanoi: Law library of Vietnam; 2017. URL: https://thuvienphapluat.vn/van-ban/the-thao-y-te/Nghi-quyet-20-NQ-TW-2017-tang-cuong-cong-tac-bao-ve-cham-soc-nang-cao-suc-khoe-nhan-dan-365599.aspx [accessed 2021-06-03]
54. Electronic Health Administration. Smart health development in the period of 2019-2025, vision to 2030. Hanoi: Ministry of Health of Vietnam; 2019. URL: https://ehealth.gov.vn/?action=News&newsId=47145 [accessed 2021-06-07]

55. Ministry of Health of Vietnam. Circular regulations on telemedicine activities. Hanoi: Legal library; 2017. URL: https://thuvienphapluat.vn/van-ban/The-thao-Y-te/Circular-49-2017-TT-BYT-on-telemedicine-374119.aspx [accessed 2020-06-07]
56. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008 May;77(5):291-304. [doi: 10.1016/j.ijmedinf.2007.09.001] [Medline: 17951106]
57. Electronic Health Administration. Circular No. 46/2018/TT-BYT regulating electronic medical records. Hanoi: Ministry of Health of Vietnam; 2018. URL: https://ehealth.gov.vn/Index.aspx?action=GioiThieu&MenuChildID=391&Id=4262 [accessed 2020-06-13]
58. Hospitals in Vietnam. Geneva: World Health Organization; 2020. URL: https://www.who.int/vietnam/health-topics/hospitals [accessed 2020-06-08]
59. Plan for implementing Electronic Health Records. Hanoi: Ministry of Health of Vietnam; 2019. URL: https://thuvienphapluat.vn/van-ban/the-thao-y-te/Quyet-dinh-5349-QD-BYT-2019-phe-duyet-Ke-hoach-trien-khai-ho-so-suc-khoe-dien-tu-428071.aspx [accessed 2020-06-13]
60. Doan EZ. Vietnam: number of Facebook users 2017-2023. Statista. 2020. URL: https://www.statista.com/statistics/553800/facebook-penetration-in-vietnam/ [accessed 2020-06-08]
61. Vietnamese people spend much of their time online. Hanoi: Vietnam News; 2018. URL: https://vietnamnews.vn/economy/481777/vietnamese-people-spend-much-of-their-time-online.html#:~:text=The%20figures%20released%20by%20the,of%203.55%20hours%20per%20day [accessed 2021-05-26]
62. Do TTT, Le MD, Van Nguyen T, Tran BX, Le HT, Nguyen HD, et al. Receptiveness and preferences of health-related smartphone applications among Vietnamese youth and young adults. BMC Public Health 2018 Jun 19;18(1):764 [FREE Full text] [doi: 10.1186/s12889-018-5641-0] [Medline: 29921258]
63. Tran BX, Zhang MW, Le HT, Nguyen HD, Nguyen LH, Nguyen QLT, et al. What drives young Vietnamese to use mobile health innovations? Implications for health communication and behavioral interventions. JMIR Mhealth Uhealth 2018 Nov 30;6(11):e194 [FREE Full text] [doi: 10.2196/mhealth.6490] [Medline: 30251704]
64. Individuals using the internet (% of population)—Vietnam. World Bank. 2021. URL: https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=VN [accessed 2021-05-31]
65. Vu LTH, Nguyen NTK, Tran HTD, Muhajarine N. mHealth information for migrants: an e-health intervention for internal migrants in Vietnam. Reprod Health 2016 May 14;13(1):55 [FREE Full text] [doi: 10.1186/s12978-016-0172-6] [Medline: 27180147]
66. Ngan TT, Do VV, Huang J, Redmon PB, Minh HV. Willingness to use and pay for smoking cessation service via text-messaging among Vietnamese adult smokers, 2017. J Subst Abuse Treat 2019 Sep;104:1-6. [doi: 10.1016/j.jsat.2019.05.014] [Medline: 31370973]
67. Tran BX, Le XTT, Nguyen PN, Le QNH, Mai HT, Nguyen HLT, et al. Feasibility of e-Health interventions on smoking cessation among Vietnamese active internet users. Int J Environ Res Public Health 2018 Dec 20;15(1):1 [FREE Full text] [doi: 10.3390/ijerph15010165] [Medline: 29361694]
68. Bates LA, Hicks JP, Walley J, Robinson E. Evaluating the impact of Marie Stopes International's digital family planning counselling application on the uptake of long-acting and permanent methods of contraception in Vietnam and Ethiopia: a study protocol for a multi-country cluster randomised controlled trial. Trials 2018 Aug 04;19(1):420 [FREE Full text] [doi: 10.1186/s13063-018-2815-0] [Medline: 30075739]
69. Nguyen QT, Naguib RNG, Abd Ghani MK, Bali RK, Lee IM. An analysis of the healthcare informatics and systems in Southeast Asia: a current perspective from seven countries. Int J Electron Healthc 2008;4(2):184-207. [doi: 10.1504/IJEH.2008.019792] [Medline: 18676343]
70. Lam JA, Dang LT, Phan NT, Trinh HT, Vu NC, Nguyen CK. Mobile health initiatives in Vietnam: scoping study. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e106 [FREE Full text] [doi: 10.2196/mhealth.8639] [Medline: 29691214]
71. McBride B, Nguyen LT, Wiljer D, Vu NC, Nguyen CK, O'Neil J. Development of a maternal, newborn and child mHealth intervention in Thai Nguyen province, Vietnam: protocol for the mMom Project. JMIR Res Protoc 2018 Jan 11;7(1):e6 [FREE Full text] [doi: 10.2196/resprot.7912] [Medline: 29326095]
72. McBride B, O'Neil JD, Hue TT, Eni R, Nguyen CV, Nguyen LT. Improving health equity for ethnic minority women in Thai Nguyen, Vietnam: qualitative results from an mHealth intervention targeting maternal and infant health service access. J Public Health (Oxf) 2018 Dec 01;40(suppl_2):ii32-ii41 [FREE Full text] [doi: 10.1093/pubmed/fdy165] [Medline: 30252117]
73. Sabin LL, Larson WA, Le BN, Herman AR, Viet NH, Albanese RR, et al. Benefits and limitations of text messages to stimulate higher learning among community providers: participants' views of an mHealth intervention to support continuing medical education in Vietnam. Glob Health Sci Pract 2017 Jun 27;5(2):261-273 [FREE Full text] [doi: 10.9745/GHSP-D-16-00348] [Medline: 28655802]
74. Larson Williams A, Hawkins A, Sabin L, Halim N, Le Ngoc B, Nguyen VH, et al. Motivating HIV providers in Vietnam to learn: a mixed-methods analysis of a mobile health continuing medical education intervention. JMIR Med Educ 2019 Apr 18;5(1):e12058 [FREE Full text] [doi: 10.2196/12058] [Medline: 30998220]

75. Gill CJ, Le NB, Halim N, Chi CTH, Nguyen VH, Bonawitz R, et al. mCME project V.2.0: randomised controlled trial of a revised SMS-based continuing medical education intervention among HIV clinicians in Vietnam. BMJ Glob Health 2018;3(1):e000632 [FREE Full text] [doi: 10.1136/bmjgh-2017-000632] [Medline: 29527350]
76. Bonawitz R, Bird L, Le NB, Nguyen VH, Halim N, Williams AL, et al. Implementing the mobile continuing medical education (mCME) project in Vietnam: making it work and sharing lessons learned. Mhealth 2019;5:7 [FREE Full text] [doi: 10.21037/mhealth.2019.02.01] [Medline: 30976599]
77. Pollack TM, Nhung VTT, Vinh DTN, Hao DT, Trang LTT, Duc PA, et al. Building HIV healthcare worker capacity through telehealth in Vietnam. BMJ Glob Health 2020;5(4):e002166 [FREE Full text] [doi: 10.1136/bmjgh-2019-002166] [Medline: 32337087]
78. Davis MC, Can DD, Pindrik J, Rocque BG, Johnston JM. Virtual interactive presence in global surgical education: international collaboration through augmented reality. World Neurosurg 2016 Mar;86:103-111 [FREE Full text] [doi: 10.1016/j.wneu.2015.08.053] [Medline: 26342783]
79. Hersh D, Hersch F, Mikuletic L, Neilson S. A Web-based approach to low-cost telemedicine. J Telemed Telecare 2003;9 Suppl 2:S24-S26. [doi: 10.1258/135763303322596174] [Medline: 14728752]
80. Imamura K, Tran TTT, Nguyen HT, Kuribayashi K, Sakuraya A, Nguyen AQ, et al. Effects of two types of smartphone-based stress management programmes on depressive and anxiety symptoms among hospital nurses in Vietnam: a protocol for three-arm randomised controlled trial. BMJ Open 2019 Apr 08;9(4):e025138 [FREE Full text] [doi: 10.1136/bmjopen-2018-025138] [Medline: 30962230]
81. GDP growth (annual %)—Vietnam. Washington: World Bank; 2020. URL: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=VN [accessed 2020-06-08]
82. Vietnam 2035: toward prosperity, creativity, equity and democracy. Washington: World Bank; 2016. URL: https://openknowledge.worldbank.org/bitstream/handle/10986/23724/9781464808241.pdf?sequence=12&isAllowed=y [accessed 2021-05-24]
83. Das K. Vietnam: growing demand for healthcare services. Hanoi: Vietnam Briefing; 2018 Sep 14. URL: https://www.vietnam-briefing.com/news/vietnam-growing-demand-healthcare-services.html/?hilite=%27Growing%27%2C%27Demand%27%2C%27Healthcare%27%2C%27Services%27 [accessed 2020-06-08]
84. The Congress of Vietnam. The law of disease examination and treatment. Hanoi: Government of the Socialist Republic of Vietnam; 2009. URL: http://vanban.chinhphu.vn/portal/page/portal/chinhphu/hethongvanban?class_id=1&_page=1&mode=detail&document_id=92348 [accessed 2020-07-25]
85. Electronic Health Administration. Decision 4376/QĐ-BYT, promulgating the regulation on the establishment, use and management of health identifications. Hanoi: Ministry of Health of Vietnam; 2019. URL: https://ehealth.gov.vn/Index.aspx?action=Detail&MenuChildID=&Id=4372 [accessed 2020-06-13]
86. The Congress of Vietnam. Social insurance Law. Hanoi: Law library; 2014. URL: https://thuvienphapluat.vn/van-ban/bao-hiem/Luat-Bao-hiem-xa-hoi-2014-259700.aspx [accessed 2020-07-25]
87. Castel P, Pick A. Increasing social insurance coverage in Viet Nam's SMEs. OECD. 2018. URL: https://www.oecd.org/dev/inclusivesocietiesanddevelopment/Increasing_social_insurance_coverage_in_Vietnam_SMES.pdf [accessed 2021-05-24]
88. Somanathan AT, Dao H, Hurt K, Fuenzalida-Puelma H. Moving toward universal coverage of social health insurance in Vietnam—assessment and options. Washington: World Bank; 2014. URL: http://documents1.worldbank.org/curated/en/383151468138892428/pdf/Moving-toward-universal-coverage-of-social-health-insurance-in-Vietnam-assessment-and-options.pdf [accessed 2021-05-24]
89. Press release on labor and employment situation in the first quarter of 2019. Hanoi: General Statistics Office of Vietnam; 2019. URL: https://www.gso.gov.vn/en/employment/ [accessed 2021-05-26]
90. Thoa D. The number of people participating in compulsory social insurance decreases and voluntary social insurance increases. Hanoi: Vietnam Communist Party electronic newspaper; 2020. URL: http://dangcongsan.vn/bao-hiem-xa-hoi-bao-hiem-y-te-vi-an-sinh-xa-hoi/tin-tuc/so-nguoi-tham-gia-bao-hiem-xa-hoi-bat-buoc-giam-bao-hiem-xa-hoi-tu-nguyen-tang-556718.html [accessed 2020-07-26]
91. Prime Minister of Vietnam. Decision 538/QD-TTg: Approving the project on implementation of program to comprehensive health insurance period 2012-2015 and toward 2020. Hanoi: Legal library; 2013. URL: https://thuvienphapluat.vn/van-ban/bao-hiem/quyet-dinh-538-qd-ttg-nam-2013-phe-duyet-de-an-thuc-hien-lo-trinh-bhyt-toan-dan-179586.aspx [accessed 2020-06-08]
92. Takashima K, Wada K, Tra TT, Smith DR. A review of Vietnam's healthcare reform through the Direction of Healthcare Activities (DOHA). Environ Health Prev Med 2017 Oct 30;22(1):74 [FREE Full text] [doi: 10.1186/s12199-017-0682-z] [Medline: 29165160]
93. National Assembly of Vietnam. Law on network information security No. 86/2015/QH13. Hanoi: Legislation web portal; 2015. URL: https://vanbanphapluat.co/law-no-86-2015-qh13-on-cyberinformation-security-2015 [accessed 2020-06-08]
94. Do N. MoH partners with FPT to build smart medical model. Hanoi: Vietnam Economic Times; 2018 Aug 16. URL: http://vneconomictimes.com/article/society/moh-partners-with-fpt-to-build-smart-medical-model [accessed 2020-07-27]
95. FPT launches hospital management solution applying 4.0 technology. FPT. 2018. URL: https://fpt.com.vn/en/ir/featured-news/fpt-launches-hospital-management-solution-applying-4-0-technology [accessed 2020-07-27]

96. Tuyet M. VNPT introduces solution to improving healthcare. Hanoi: Vietnam Economics Times; 2015 Jun 26. URL: http://www.vneconomictimes.com/article/business/vnpt-introduces-solution-to-improving-healthcare [accessed 2020-07-27]
97. Schneider H, Hill S, Blandford A. Patients know best: qualitative study on how families use patient-controlled personal health records. J Med Internet Res 2016 Feb 24;18(2):e43 [FREE Full text] [doi: 10.2196/jmir.4652] [Medline: 26912201]
98. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
99. Bate P, Robert G. Bringing User Experience to Healthcare Improvement: The Concepts, Methods and Practices of Experience-based Design. Oxford: Radcliffe Publishing; 2007.
100. Steen M, Manschot M, de Koning KN. Benefits of co-design in service design projects. Int J Des. 2: 5; 2011. URL: http://www.ijdesign.org/index.php/IJDesign/article/view/890/346 [accessed 2021-05-24]
101. Fourth industrial revolution. World Economic Forum. 2020. URL: https://www.weforum.org/focus/fourth-industrial-revolution [accessed 2020-06-14]

Abbreviations

BHS: Bespoke Health Care System
EHR: electronic health record
EMR: electronic medical record
HIS: health information system
MOH: Ministry of Health of Vietnam
NCD: noncommunicable disease
OPP: out-of-pocket payment
SWOT: strength, weakness, opportunity, and threat analysis
THE: total health expenditure
WHO: World Health Organization

Edited by G Eysenbach; submitted 27.09.20; peer-reviewed by D Adriaans, N Phan; comments to author 05.11.20; revised version received 29.12.20; accepted 14.04.21; published 04.06.21.

Please cite as:
Dang TH, Nguyen TA, Hoang Van M, Santin O, Tran OMT, Schofield P
Patient-Centered Care: Transforming the Health Care System in Vietnam With Support of Digital Health Technology
J Med Internet Res 2021;23(6):e24601
URL: https://www.jmir.org/2021/6/e24601
doi:10.2196/24601
PMID:34085939

Viewpoint

Development, Status Quo, and Challenges to China’s Health Informatization During COVID-19: Evaluation and Recommendations

Mian Huang[1,2], MSc; Jian Wang[1,2], Prof Dr; Stephen Nicholas[3,4,5,6,7], Prof Dr; Elizabeth Maitland[8], Prof Dr; Ziyue Guo[1,2], BEc

[1]Dong Fureng Institute of Economic and Social Development, Wuhan University, Beijing, China
[2]Center for Health Economics and Management, School of Economics and Management, Wuhan University, Wuhan, China
[3]Australian National Institute of Management and Commerce, Sydney, Australia
[4]Newcastle Business School, University of Newcastle, Newcastle, Australia
[5]Guangdong Institute for International Strategies, Guangdong University of Foreign Studies, Guangzhou, China
[6]School of Economics, Tianjin Normal University, Tianjin, China
[7]School of Management, Tianjin Normal University, Tianjin, China
[8]School of Management, University of Liverpool, Liverpool, United Kingdom

Corresponding Author:
Jian Wang, Prof Dr
Dong Fureng Institute of Economic and Social Development
Wuhan University
No 54 Dongsi Lishi Hutong
Dongcheng District
Beijing, 100010
China
Phone: 86 13864157135
Email: wangjian993@whu.edu.cn

Abstract

By applying advanced health information technology to the health care field, health informatization helps optimize health resource allocation, improve health care services, and realize universal health coverage. COVID-19 has tested the status quo of China’s health informatization, revealing challenges to the health care system. This viewpoint evaluates the development, status quo, and practice of China’s health informatization, especially during COVID-19, and makes recommendations to address the health informatization challenges. We collected, assessed, and evaluated data on the development of China’s health informatization from five perspectives—health information infrastructure, information technology (IT) applications, financial and intellectual investment, health resource allocation, and standard system—and discussed the status quo of the internet plus health care service pattern during COVID-19. The main data sources included China’s policy documents and national plans on health informatization, commercial and public welfare sources and websites, public reports, institutional reports, and academic papers. In particular, we extracted data from the 2019 National Health Informatization Survey released by the National Health Commission in China. We found that China developed its health information infrastructure and IT applications, made significant financial and intellectual informatization investments, and improved health resource allocations. Tested during COVID-19, China’s current health informatization system, especially the internet plus health care system, has played a crucial role in monitoring and controlling the pandemic and allocating medical resources. However, an uneven distribution of health resources and insufficient financial and intellectual investment continue to challenge China’s health informatization. China’s rapid development of health informatization played a crucial role during COVID-19, providing a reference point for global pandemic prevention and control. To further promote health informatization, China’s health informatization needs to strengthen top-level design, increase investment and training, upgrade the health infrastructure and IT applications, and improve internet plus health care services.

(J Med Internet Res 2021;23(6):e27345) doi:10.2196/27345

KEYWORDS

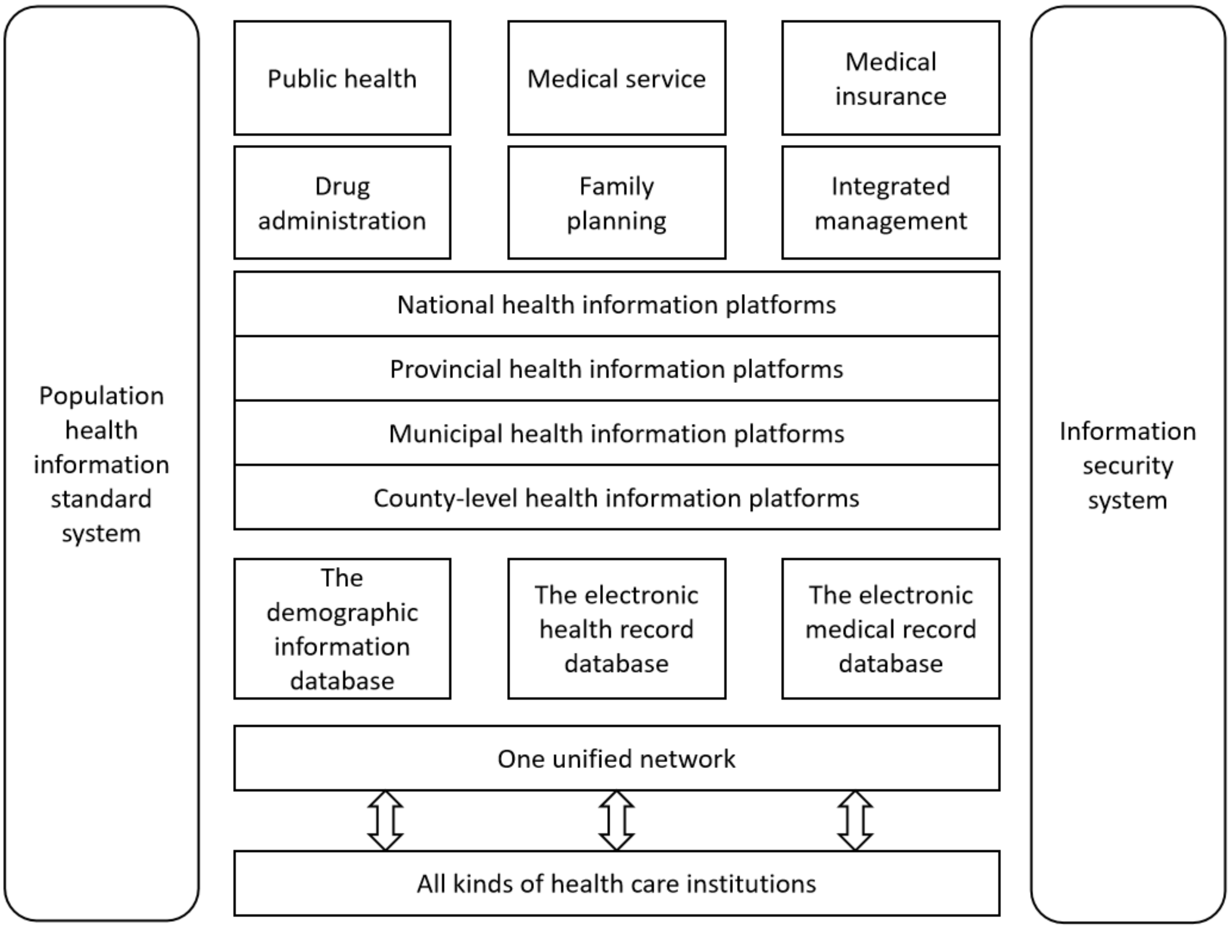
health informatization; COVID-19; health policy; digital health; health information technology; China

Introduction

Background

Building on China’s highly developed information-based society, including big data, cloud computing, mobile internet, and artificial intelligence, health information technology (HIT) provides a key impetus for the Chinese government to maximize health resource allocation, address the uneven geographical distribution of medical resources, ensure China’s universal health coverage, and enable health care providers to optimize health care services with lower medical costs and better quality [1-6]. China’s health informatization has developed through several stages, and the policy system at every stage has sought consistency and coordination. In 2010, the National Health and Family Planning Commission (NHFPC) issued the “3521 framework” for health informatization, which was then upgraded to the “4631-2 framework” in 2013 [7-10]. Figure 1 presents a structural figure of the “4631-2 framework” based on NHFPC policy documents. As shown in Figure 1, the “4631-2 framework” referred to 4 levels of national health information platforms (national, provincial, municipal, and county level), 6 primary applications (public health, medical service, medical insurance, drug administration, family planning, and integrated management), 3 major health information databases (the demographic information database [DID], the electronic health record database [EHRD], and the electronic medical record database [EMRD]), 1 unified network covering all kinds of health care institutions (including traditional Chinese medical institutions), and 2 systems (population health information standard system and information security system) [8,11].

Figure 1. The “4631-2 framework” of health informatization.



Building on the “4631-2 framework” of health informatization, the 2016 Plan of Healthy China 2030 committed to expanding the health information service system. The NHFPC released the National Population Health Informatization Development Plan in 2017, emphasizing the essential role of health informatization in national IT (information technology) construction, health reform and the Healthy China initiative [12]. In 2018, the National Health Commission (NHC) was established, further emphasizing the role of health informatization by targeting comprehensive and lifelong intelligent health services by utilizing advanced technologies.

Objective

Surprisingly, the development of China’s health informatization system has not been detailed nor its status quo effectiveness assessed. The early COVID-19 period, where, for example, the extensive use of internet plus health care, including internet hospitals and telemedicine, helped to ensure the normal operation of the medical system, provides a unique opportunity to evaluate China’s health informatization [13-15]. Our study aimed to map the development of China’s health informatization; evaluate China’s status quo health informatization experiences and lessons learned during

COVID-19; identify the challenges to health informatization; and provide recommendations for the improvement of health informatization.

Data Sources and Approach

The NHC conducted a 2018 survey on national health informatization, which was published in the 2019 National Health Informatization Survey Report [16]. The NHC also held several expert meetings to assess the survey data, which was also published in the survey report [16]. The data in the NHC Survey Report can be divided into two parts: regional health informatization and hospital informatization. The Regional Health Informatization Survey covered 1650 health commissions (HCs), including all 32 provincial HCs, 69.6% (296/425) of municipal HCs, and 47.1% (1322/2803) of county-level HCs. The Hospital Informatization Survey covered 5470 hospitals, including 70.1% (1480/2112) of tertiary hospitals, 62.1% (3730/6004) of tertiary hospitals, and 6.2% (260/4180) of other hospitals (at lower levels). As shown in Multimedia Appendix 1, the survey identified key aspects of health informatization, including infrastructure, IT application, capital investment, and human resources, providing data for evaluating the status quo of China’s health informatization. Multimedia Appendix 1 also displays other information we collected, interpreted, and evaluated from primary sources, academic papers, government reports, and institutional reports regarding the current status and challenges of health informatization.

Figure 1 sets out a diagrammatic representation of the conceptual basis of China’s health informatization, while Multimedia Appendix 1 summarizes the main primary and secondary sources used to outline the development, evaluate the status quo, and identify the challenges to China’s health informatization.

Development and Status Quo of China’s Health Informatization

Health Information Infrastructure

China’s health information platforms form the basis of the country’s regional health informatization. In 2012, the Ministry of Health in China proposed a 3-level (national, provincial, and regional) health information platform [9]. As displayed in Figure 1, the health information platform was extended in 2013 to consist of 4 administrative level platforms: the national, provincial, municipal, and county-level information platforms [10]. On each platform, the medical data from different regions were integrated and shared [7]. The platforms were designed to be unified, authoritative, and interconnected so as to exploit and statistically analyze health information of the whole population, thus providing support for management and decision making on a national health level [12,17,18]. The platforms also provided the basis for an information business system covering the whole industry chain of health and medical big data, linking medical care, medical insurance, and medicine. By 2018, 93.8% (30/32) of provincial health information platforms, 66.2% (196/296) of municipal health information platforms, and 48.2% (637/1322) of county-level health information platforms had been established [16]. During COVID-19, by sharing data with internet hospitals and health care institutions at all levels, regional health information platforms connected online and offline services to provide patients with full process services before, during, and after diagnosis [19,20]. During the COVID-19 period, the NHC promoted new infrastructure construction in health informatization, emphasizing the need to improve health information platforms and health information databases [19].

National health information databases include three major databases: the DID, the EHRD, and the EMRD. The DID contains fundamental population information, family planning service management information, and the nonresident population management information. The EMRD stores all information from electronic medical records while the EHRD holds residents’ personal health information from the electronic health record (EHR), which is defined as “digitally stored health care information about an individual’s lifetime with the purpose of supporting continuity of care, education, and research and ensuring confidentiality at all times” [7,21]. In recent years, the NHC has promoted the EHR to be opened to individuals. In pilots undertaken in selected areas, residents are able to enter the EHR system with a password and access their own EHRs [22,23]. The 2010 “3521 framework” contained only two databases—the EHRD and the EMRD [24]—with DID added under the 2013 “4631-2 framework” [25].

The three major databases are relatively independent but interrelated. Supported by the national health information platform, information in the three databases is shared and dynamically updated to ensure consistency, accuracy, and integrity of the information [12,26]. According to the 2019 National Health Informatization Survey, the average construction rate of EHRDs was 79.5% (1312 out of 1650 HCs had built EHRDs), the highest among the three databases, with 93.8% (30/32) of provincial, 84.8% (251/296) of municipal, and 78.0% (1031/1322) of county-level EHRDs. The average construction rate of the DIDs was 77.5% (1278 out of 1650 HCs), with 100% (32/32) of provincial, 80.4% (238/296) of municipal, and 76.2% (1008/1322) of county-level DIDs. Lowest among these three databases, the average construction rate of EMRD was 49.5% (816 out of 1650 HCs), which was 93.8% (30/32) of provincial, 59.5% (176/296) of municipal, and 46.1% (610/1322) of county-level EMRDs [16].

IT Applications

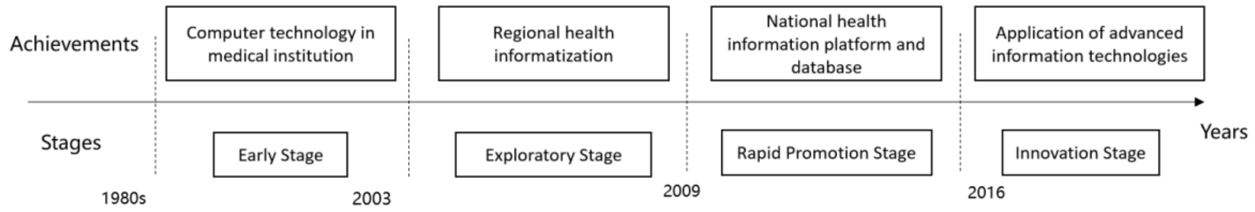
To illustrate the historical development of technology applications, we constructed Figure 2, which sets out the four development stages in China’s health informatization: the early stage from the 1980s to 2003, the exploratory stage from 2003 to 2009, the rapid development stage from 2009 to 2016, and the innovation stage since 2016 [27,28]. In the early stage (1980s-2003), the emphasis was placed on the application of computer technology in information systems, such as the financial management system and charging system, in large medical institutions. Manual operations were replaced by computer functions in the management of traditional businesses, including financial management, fee management, and drug management. At the exploratory stage (2003-2009), drawing on the experience of SARS (severe acute respiratory syndrome) in 2003, China strengthened its informatization in health emergency command, disease prevention and control, and public health resources and health information management [29]. Importantly, a direct internet reporting system for infectious diseases and public health emergencies was established, and regional health informatization emerged. In the context of health care reform, the rapid development stage (2009-2016) completed the construction of the national health information platform and database, dealing with the information islands in health services as well as realizing the platform connectivity and information sharing of the medical and health systems. Since 2016, more advanced information technologies, such as internet plus, big data, cloud computing and artificial intelligence have been widely applied to health industries. In the innovation stage, a people-centered medical service principle was adopted [30], with advanced IT applications improving the health information system and providing more diverse health information services [12].

Currently, advanced technologies commonly used in China's health informatization include cloud computing, big data, Internet of Things, mobile internet, and artificial intelligence. The 2019 National Health Informatization Survey showed that advanced technologies, especially mobile internet and big data, were more commonly applied on provincial than municipal and county platforms. In 2018, on provincial health information platforms, the application rate of cloud computing was 68.8% (22 out of 32 provinces), big data was 53.1% (17 out of 32 provinces), Internet of Things was 21.9% (7 out of 32 provinces), mobile internet was 71.9% (23 out of 32 provinces), and artificial intelligence was 12.5% (4 out of 32 provinces). On municipal and county-level platforms, advanced technologies were less utilized, where 40.2% (119/296) of municipal and 46.1% (609/1322) of county-level platforms had not utilized any advanced technologies [16]. Higher-level hospitals had higher IT application rates, with mobile internet being the most commonly used technology (tertiary hospitals: 747/1480, 50.5%; secondary hospitals: 767/3730, 20.6%; and other hospitals: 63/260, 24.2%) and artificial intelligence being the least used technology (tertiary hospitals: 137/1480, 9.3%; secondary hospitals: 62/3730, 1.7%; and other hospitals: 3/260, 1.2%).

COVID-19 tested the informatization of China's health system. In COVID-19 prevention and control, advanced information technologies were effectively utilized in monitoring and forecasting the pandemic's trends as well as constraining the spread of the virus. Based on big data, the intelligent risk assessment system and automatic early warning system were utilized to fight the virus. Big data was used to track the mobility of the population and locate crowds after the outbreak. With the help of artificial intelligence, the number of infections was estimated, and high-risk areas identified, which helped the government to implement timely control methods and determine the allocation of resources [31]. Using a series of data statistics and model analysis, big data population movement information on Baidu Maps was one of the most popular technologies to estimate the number and locations of patients with COVID-19 and carriers. The big data methods also predicted disease trends based on real-time reporting of new cases.

In order to improve data utilization, some hospitals promoted the connection between hospital information systems and regional health information platforms. Hospital information departments extracted data from major information systems such as Hospital Information System, Laboratory Information System, Picture Archiving and Communication Systems, and EMRs to build a clinical data center [32,33]. An induction model based on artificial intelligence was constructed to help predict the risk and trend of infectious diseases and enhance prevention efforts [1].

Figure 2. The development history of health informatization in China.



Financial and Intellectual Investment

As seen in Multimedia Appendix 1, the importance of health informatization can be assessed by China's increased levels of investment. During the 2011-2015 12th Five-Year Plan, the central government invested more than ¥10 billion (US $1.5 billion) into health informatization [7]. In 2017, the first phase of the National Health Security Informatization Project was approved by the National Development and Reform Commission, with ¥340 million (US $52.5 million) invested in national platform construction and 32 national hospital informatization projects approved by the NHFPC, and ¥310 million (US $47.9 million) invested by the NHFPC in the construction of information platforms and clinical data centers [34]. In 2018, the Ministry of Finance invested ¥670 million (US $103.5 million) to equip 832 poor county hospitals and 1664 primary-level medical institutions with telemedicine equipment. In 2019, the central government allocated ¥180 million (US $27.8 million) to increase technology applications, including the internet, big data, and artificial intelligence in 9 provinces, with the aim of improving the diagnosis and treatment capacity of family doctors [35].

Although governmental investment in health informatization expanded rapidly, it failed to cover China's current financial informatization needs, reflecting a lack of stable and sustainable capital input. For example, China's health expenditure accounted for 7.56% of the financial expenditure, while health informatization expenditures accounted for only 0.7% of the total health expenditure [36]. Second, there were obvious regional differences in governmental investment, with Beijing, Guangzhou, Shanghai, and the eastern coastal provinces achieving rapid development of health informatization versus the less developed provinces [37]. The capital input in informatization was positively associated with the local economic development level, where informatization tended to increase regional health inequality [16]. The 2019 National Health Informatization Survey showed that hospital informatization investment was limited. The proportion of informatization investment in a hospital's annual revenue was between 0.1% and 1% in 67.3% (3681/5470) of hospitals and 1% to 5% in 23.7% (1296/5470) of hospitals, with only 3.7% (202/5470) of hospitals investing over 5% of their annual revenue in informatization. With the construction of health informatization requiring multisectoral collaboration, the lack of social supervision led to a lack of initiative and a waste of resources. In 2017, according to the China Hospital Information Management Association survey, hospital self-raised funds accounted for 68.8% of the total investment in hospital informatization, with government financial funds accounting for 28.4% and other funds accounting for 2.8% of the total investment [38,39]. Although the Chinese government has implemented policies to encourage enterprise initiative, an incentive mechanism to encourage social capital to participate in health informatization construction has not been set up in China. Measures to promote government-enterprise cooperation in health informatization are still being explored [12,40].

Both the quantity and quality of technical informatization personnel failed to meet demand [12]. The number of health informatization personnel was small. In 2019, the average staff number in provincial information departments was only 20, while 91.6% (271/296) of municipal information departments and 93.6% (1238/1322) of county-level information departments had less than 9 informatization staff. In hospitals, the information department of tertiary hospitals had on average 10 staff, while 74.1% (2764/3730) of secondary hospitals and 89.5% (233/260) of other hospitals had less than 4 informatization staff. In particular, there was a lack of highly educated and specialized informatization personnel and an uneven distribution of high-quality personnel in administration and hospitals at different levels [41]. High-level administrations and hospitals tended to have staff with higher education [16], with the proportion of staff with a bachelor's degree or above in provincial information departments (88.9%) and in municipal information departments (83.8%) significantly higher than that in county-level information departments (51.6%). In tertiary hospitals, 67.5% of information department staff had bachelor's degrees, while in secondary and other hospitals, most staff had degrees below the junior college level. The proportion of staff with a master's degree or above in tertiary hospitals (14.3%) was significantly higher than that in secondary hospitals (2.3%) and in other hospitals (0.7%). Staff specialized in computer science accounted for 85.1% of informatization staff in tertiary hospitals compared to 49.1% in secondary and 26.5% in other hospitals. In addition, IT staff with senior professional titles were scarce in all types of hospitals, with only 2.0% in tertiary, 1.9% in secondary, and 0.6% in other hospitals [16].

Health Resource Allocation

Medical resources distribution and health service capacity were unevenly distributed regionally and by type of medical institution [3,4]. The data from Multimedia Appendix 1 reveals that the construction strategies of health informatization varied according to the health care resource allocation status in different regions [1].

For developed regions (mainly the eastern coast) with abundant medical resources, the local health administration undertook the major responsibilities for integrating high-quality medical resources and promoting health informatization. These regions had stronger abilities to independently develop informatization and provide stable and sufficient funds for informatization. Bottom-up building of health information platforms and data centers were established by the local health administration to connect the information system of health institutions at all levels in the administrative region, achieving the sharing and exchange of medical information. For example, municipal and county-level health commissions in each province independently carried out informatization construction, driving the development of provincial health informatization and promoting the development of regional health information infrastructures, information technologies, and professionals. Municipal and county-level health resource information was integrated, and the connectivity of health information within the province was realized [24].

In developing regions, mainly in western and central China, where medical resources were scarce, large hospitals generally connected with several local primary health care institutions to realize medical information exchange and sharing [24]. For instance, the West China Hospital of Sichuan University cooperated with surrounding secondary hospitals and community health service centers to achieve regional medical collaboration. Shengjing Hospital of China Medical University, in cooperation with public hospitals at all levels in Liaoning and Shenyang, constructed the Shengjing Medical Alliance, a nonprofit medical cooperative alliance. In these less developed and health resource scarce regions, unified management was implemented within the medical alliances to integrate medical resources at different levels. To optimize medical resources, resource-poor regions utilized telemedicine, or the use of telecommunications technologies to provide medical information and services, [1,4,42] for remote consultation and two-way referral, where primary medical units can get support from large hospitals and gain access to high-quality medical resources and services.

The COVID-19 pandemic typically led to a surge in demand for medical care, which overwhelmed local capabilities [43,44]. During the pandemic, telemedicine played an increasingly important role in medical resource allocation [1,4,42]. As the worst-hit area of COVID-19 in China, Hubei Province's medical institutions were overloaded with a severe shortage of medical resources and services [43]. Due to the implementation of control measures such as lockdowns, traffic control, and community closure management, residents' access to health care resources was severely restricted. Telemedicine was widely used to overcome geographical barriers and expand the supply of medical resources [45-47]. With the aid of informatization technologies, many doctors from other parts of China provided remote diagnosis and treatment services. For instance, the medical team from Guangdong Province built an internet hospital to provide online clinical support for patients in Jingzhou, Hubei Province. More than 1300 doctors from 15 medical institutions in Guangdong participated, with the total number of online visitors exceeding 100,000 within 18 days of its launch, making up for local medical service deficits in Hubei Province [48].

The Standard System

As seen in Figure 1 and Multimedia Appendix 1, emphasis has been placed on the standardization of health information, or the standard system [12]. In 2006, the Special Committee on Health Information Standards was approved by the Ministry of Health, where work on health information norms were standardized. Started in 2012, the National Medical Health Information System's Interconnection and Interoperability Standardization Certification [16] aimed to strengthen the management and promote the implementation of health information standards, so as to achieve health information interoperability [11]. The Regional Health Informatization Survey and Hospital Informatization Survey found that by 2017, 54 regional platforms (1 provincial platform, 45 municipal platforms, and 8 county-level platforms), and 90 hospitals (84 tertiary hospitals and 6 secondary hospitals) had been certified. Between 2018 and 2019, the number certified significantly increased with an additional 48 regional platforms and 101 hospitals certified [49].

The security system of regional health information platforms contained 4 levels: physical security, network security, system security, and application security. Physical security required hardware security and the geographical location security of the computer room and key infrastructure. Network security included link redundancy, firewall setup, and intrusion detection. System security included data backup, virus prevention, operating system security, and vulnerability checks. Application security included identity authentication and authority management [8,50]. By 2018, all provincial platforms, 90.9% (269/296) of municipal platforms, and 78.9% (1043/1322) of county-level platforms had formulated an information security system. Meanwhile, 99.2% (1468/1480) of tertiary hospitals, 95.9% (3577/3730) of secondary hospitals, and 81.5% (212/260) of other hospitals had formulated an information security system. By August 2020, with 227 effective information standards approved, the standard system for regional health informatization and hospital informatization had been basically established throughout the health system [26].

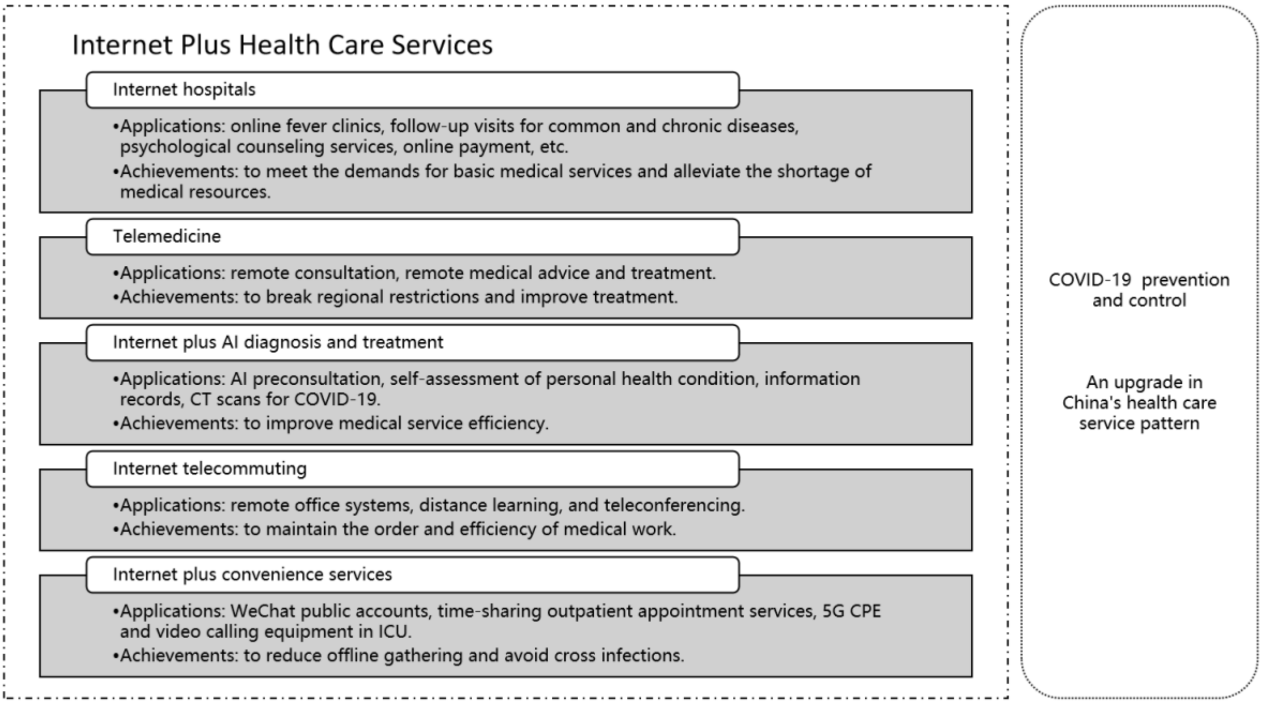
During COVID-19, continuing shortcomings in the standard system were addressed by the 2020 Notice on Strengthening the National Health Information Standardization System, which defined four key tasks for the construction of the national health information standardization system: (1) to promote the standardization of the national health information infrastructure, including the national health information platform, national hospital information platform, government services integration platform, and to standardize the information from grassroots health institutions, public health systems, and traditional Chinese medicine hospitals; (2) to strengthen the standardization of the DID, the EHRD, the EMRD, and the health care resource database; (3) to promote the standardization of advanced IT applications, comprising five technologies: internet plus health care, health and medical big data, health care artificial intelligence, health care 5G technology, and blockchain technology for health care; and (4) to improve the standardization of network security, namely industry network security standard system, data security standard, and industry application security standard. In order to improve the disease prevention and control system and enhance the ability to respond to public health emergencies, the NHC formulated the National Public Health Informatization Construction Standard and Specification on December 11, 2020, to further regulate national public health informatization, specifying 21 first-level indicators, 125 second-level indicators, and 421 third-level indicators of health service management and IT application content in all types of health institutions [26].

Even with these reforms, shortcomings in China's health information standard system remain. There was a lack of standard evaluation, and the consistency and interoperability of the standard system needed strengthening [1]. The formulation of China's standard system lagged behind the development of new technologies. The safety standards, data standards, functional standards, quality assessment standards for information technology products, and management standards for data quality were incomplete. In terms of information security standards, the ownership and usage rights of patients' health information, as well as the division of rights and responsibilities among all parties, needed further clarification in laws and regulations. There was a lack of security standard evaluation. The Regional Health Informatization Survey reported that in 2018, 75.0% (24/32) of provincial platforms, 30.4% (90/296) of municipal platforms, and 17.7% (234/1322) of county-level platforms had implemented the evaluation of classified information security protection. Only 46.4% (687/1480) of tertiary hospitals, 18.6% (694/3730) of secondary hospitals, and 10% (26/260) of other hospitals had participated in the evaluation. Although China continues to develop health information standards, due to the large number of departments involved, many standards and norms were not unified. There were also problems related to unclear management responsibility for the standards and ineffective supervision mechanism [11].

Status Quo of the Internet Plus Health Care Service Pattern During COVID-19

As shown in Figure 3, the new service pattern of internet plus health care sought to meet increasing medical demands and virus prevention and control needs during the COVID-19 pandemic.

Figure 3. The applications and achievements of internet plus health care in COVID-19. AI: artificial intelligence, CT: computed tomography, CPE: customer premise equipment, ICU: intensive care unit.



Internet plus health care has the characteristics of remoteness, efficiency, intelligence, and convenience, providing online consultation, remote diagnosis, psychological counseling, chronic disease follow-up, and drug distribution to meet people's medical demands [40,48]. During COVID-19, internet plus health care aimed to break through time and space constraints to both identify and service infected people, block transmission routes, and reduce the risk of population infection. The popularization of internet plus health care has not only improved the efficiency of health care institutions at all levels [51] but also stimulated innovation and accelerated the upgrading of China's health care service pattern. We set out below the contours of internet plus health care during the COVID-19 pandemic.

Internet Hospitals: To Meet the Demands for Basic Medical Services and Alleviate the Shortage of Medical Resources During COVID-19

Figure 3 identifies the role of internet hospitals and online fever clinics after the COVID-19 outbreak, which were widely established to provide follow-up visits for common and chronic diseases, as well as for psychological counseling services. In an internet hospital, a full set of medical services such as registration, consultation, diagnosis, and prescription were completed by means of video, voice, and graphic communication [52]. Hospitals arranged the delivery of drugs while patients made online payments [53]. By effectively serving a large number of patients with COVID-19, internet hospitals reduced diagnosis and treatment pressure in brick-and-mortar hospitals, greatly alleviating the shortage of medical resources in severely affected COVID-19 areas.

Telemedicine: To Break Regional Restrictions and Improve Treatment During COVID-19

As specified in Figure 3, remote consultation through the telemedicine platform integrated high-quality medical resources and provided support from large institutions and well-known experts to improve treatment. Due to the uneven distribution of medical resources in different regions, China attached great importance to constructing a telemedicine system, which played a particularly significant role during the pandemic. As discussed above, a typical case was the establishment of 9 telemedicine centers in Guangdong Province to support the hard-hit COVID-19 city of Jingzhou. While on-site diagnosis was carried out, remote medical advice and treatment were provided via the telemedicine platform at the same time.

Internet Plus Artificial Intelligence Diagnosis and Treatment: To Improve Medical Service Efficiency During COVID-19

As shown in Figure 3, online artificial intelligence preconsultation functions were widely used to assess COVID-19 risks and provide emotional care. People made a self-assessment of their personal health condition with the help of technologies powered by artificial intelligence, where patients provided their personal information, past medical history, and current medical conditions through dialogue before seeing a doctor. This information was recorded by artificial intelligence and integrated into the health information system, then directly added into the patient's medical record to facilitate subsequent diagnosis. Artificial intelligence was also applied to CT (computed tomography) scans for COVID-19, improving the accuracy and efficiency of front-line medical work [31].

Internet Telecommuting: To Maintain the Order and Efficiency of Medical Work During COVID-19

As noted in Figure 3, distance learning and teleconferencing had become mainstream working modes during the pandemic. Many hospitals used virtual private network (VPN) technology to establish secure data links, and combined with virtual desktop infrastructure technology, a remote office environment was built. By opening the VPN account and private cloud desktop, the remote office system enabled medical staff to log into intranet workstations to work remotely [54].

To enhance the service ability of medical staff, distance education was also used widely during the pandemic. Learning content was issued on online learning platforms. To check the effectiveness of these learning platforms, medical staff were required to take regular online tests, and test results were recorded in real time. Moreover, IT companies such as Tencent and Alibaba launched free online meeting systems to support antipandemic work, enabling the pandemic prevention working groups to arrange work, and discuss diagnosis methods and pandemic prevention countermeasures, through video conferencing. It not only reduced crowd gathering but also improved hospital management efficiency.

Internet Plus Convenience Services: To Reduce Offline Gathering and Avoid Cross Infections During COVID-19

During the pandemic, many hospitals provided internet plus convenience services for the public, as specified in Figure 3. Through the publication function of WeChat public accounts, hospitals launched real-time news and science articles about COVID-19 as part of information campaigns, which prevented rumors and helped the public acquire accurate pandemic knowledge [14,43]. In addition, many hospital information departments opened time-sharing outpatient appointment services, where the hospital's WeChat public account allowed patients to enter the appointment and registration system according to patients' personal needs. The appointment service helped separate patients' arrival time and reduced the number of people in the doctor's waiting room at one time, containing the spread of the virus [53]. To facilitate communication between intensive care unit (ICU) patients and their families, while avoiding staff gathering, some hospitals installed 5G customer premise equipment and video calling equipment in ICUs, substituting 5G mobile video for physical visits during COVID-19.

China's internet plus service pattern also played an important role for overseas Chinese people, providing authoritative and scientific knowledge of COVID-19 prevention. The platform was launched on April 7, 2020, bringing together 26 well-known Chinese medical institutions and third-party service institutions to provide free internet consulting services for overseas Chinese individuals. On April 30, a special column aimed at overseas students was launched to provide free health consultation and psychological counseling services. By September 2020, the platform had been visited more than 150 million times, with more than 6 million people providing direct consulting services [48].

Challenges and Recommendations

Challenges Ahead

According to the findings presented above, significant challenges to China's health informatization remain. Table 1 classifies and summarizes the challenges to health informatization. First, the number of informatization professionals is small, and the distribution of human resources is unbalanced. Second, the investment from government and medical institutions is insufficient to support the development of health informatization. An incentive mechanism to encourage social capital to participate in health informatization construction has not been set up in China [3,28]. Third, the IT application on health information platforms and in hospitals at all levels remains low. Fourth, the information standards and security systems are incomplete.

Table 1. Challenges to national health informatization.

Factors	Descriptions
Personnel	
Quantity	The number of senior technical personnel and grassroots practical personnel is insufficient
Structure	Human resource allocation as measured by quality, educational background, professional title, and age is unbalanced
Funds	
Financial	Financial support is generally insufficient and shows regional differences
Institutional	Low proportion of hospitals' annual revenue allotted to informatization investment
Social	A lack of social capital's participation and government-enterprise cooperation
IT[a] applications	
Regional health information platform	Low application rate on municipal and county-level information platforms
Hospital information system	Low application rate in secondary hospitals and primary hospitals
Systems	
Information standard system	A lack of system interoperability and interconnection, standard management, and supervision
Information security system	A lack of complete and specific systems for standard evaluation and health information security protection

[a] IT: information technology.

Recommendations

Below we present recommendations to further promote health informatization in China.

Strengthen Top-Level Design

As identified in Table 1, there is a need to strengthen top-level design. Legal restrictions on some key aspects of national health informatization, such as health care big data management, internet medical information services, privacy protection, information facilities security, and network security, need to be enhanced. The information standard and security systems should be designed in more unified and specific ways, and information security supervision needs to be strengthened [55]. An incentive mechanism should be set up to encourage a diverse range of organizations and institutions, including health care institutions, research institutes, universities, associations, and enterprises, to participate in the construction of health informatization.

Increase Financial Investment

As specified in Table 1, finance, taxation, and investment arrangements need to be increased to support health informatization, especially from the central government and especially in regions with limited financial resources for health informatization. To address informatization financial constraints, public-private partnerships should be expanded to strengthen the cooperation between the government, business enterprises, and various social institutions [28].

Increase Qualified Personnel

As shown in Table 1, there is an urgent need to increase qualified personnel specializing in HIT. An integrated and comprehensive training system should be built to cover junior college, undergraduate, graduate, and continuing education students to enlarge the scale of skilled workers. An incentive mechanism should be established to attract experienced professionals and maintain existing personnel in health information departments. To improve the professional level and comprehensive quality of health information personnel, measures should include establishing a national health big data research institute jointly with universities and research institutes, setting up health information majors in colleges and universities and carrying out continuing education to combine teaching content with practical work.

Promote IT Applications

Addressing the deficits in information platforms and systems identified in Table 1 requires the reallocation of health resources and promotion of advanced technologies to support the new service pattern of internet plus health care. Advanced IT applications, such as cloud computing, big data, mobile internet, and artificial intelligence, need to be popularized in grassroots health units. It is necessary to construct more internet hospitals, building an integrated online and offline medical service model covering prediagnosis, in-treatment, and postdiagnosis to improve health care service efficiency.

Conclusions

As shown by its crucial role in COVID-19 prevention and control, health informatization was "fit for purpose" in China. Developed over the past decade, health information platforms and health information databases facilitated data sharing between health care institutions at all levels, connecting online and offline services. Advanced information technologies, including cloud computing, big data, and artificial intelligence, effectively monitored and forecasted the pandemic's trend that facilitated health authorities' COVID-19 control and prevention ability. The new service pattern of internet plus health care broke through time and space constraints, reduced cross infections, improved medical service quality and efficiency [51], alleviated the shortage of medical resources, and met the increasing medical demands during COVID-19. However, China’s health informatization is a work in progress; informatization is incomplete and the uneven regional distribution of health resources continues to persist. To address the challenges to health informatization, we recommend the strengthening of top-level design, increasing investment and training, upgrading the health infrastructure and IT applications, and improving internet plus health care services.

Acknowledgments

The authors appreciate the support from the National Health Commission of the People’s Republic of China. The authors also thank the anonymous reviewers and the editor for their constructive comments.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data sources on the development and status quo of China’s health informatization from government, commercial, and public welfare sources and websites; academic papers; public reports; institutional reports; and fieldwork.

[DOCX File , 18 KB - jmir_v23i6e27345_app1.docx]

References

1. Liang J, Zheng X, Chen Z, Dai S, Xu J, Ye H, et al. The experience and challenges of healthcare-reform-driven medical consortia and Regional Health Information Technologies in China: A longitudinal study. Int J Med Inform 2019 Nov;131:103954 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103954] [Medline: 31513943]
2. Li L, Fu H. China's health care system reform: Progress and prospects. Int J Health Plann Manage 2017 Jul;32(3):240-253 [FREE Full text] [doi: 10.1002/hpm.2424] [Medline: 28612498]
3. Yao H, Zhan C, Sha X. Current situation and distribution equality of public health resource in China. Arch Public Health 2020 Sep;78:86 [FREE Full text] [doi: 10.1186/s13690-020-00474-3] [Medline: 32983449]
4. Central Committee and the State Council, People's Republic of China. Opinions on Deepening the Reform of the Health Care System. 2009 Mar 17. URL: http://www.gov.cn/gongbao/content/2009/content_1284372.htm [accessed 2020-10-30]
5. Liu G, Chen Y, Qin X. Transforming rural health care through information technology: an interventional study in China. Health Policy Plan 2014 Dec;29(8):975-985. [doi: 10.1093/heapol/czt078] [Medline: 24179153]
6. Buntin MB, Jain SH, Blumenthal D. Health information technology: laying the infrastructure for national health reform. Health Aff (Millwood) 2010 Jun;29(6):1214-1219 [FREE Full text] [doi: 10.1377/hlthaff.2010.0503] [Medline: 20530358]
7. Li C, Xu X, Zhou G, He K, Qi T, Zhang W, et al. Implementation of National Health Informatization in China: Survey About the Status Quo. JMIR Med Inform 2019 Feb 21;7(1):e12238 [FREE Full text] [doi: 10.2196/12238] [Medline: 30789350]
8. Chen M. Theory and Method of Regional Population Health Informatization. Beijing, China: Science Press; 2016.
9. National Health Commission, People's Republic of China. Guiding Opinions on Strengthening Health Informatization Construction. 2021 Jun 15. URL: http://www.nhc.gov.cn/zwgk/wtwj/201304/e1b9fd5596ce4a5e8123337552358b38.shtml [accessed 2020-10-30]
10. National Health Commission, People's Republic of China. Guiding Opinions on Accelerating the Information Construction of Population Health. 2013 Dec 11. URL: http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=ffa4b17e89f6414d9b8a98e5b07f47b7 [accessed 2020-10-30]
11. Hu J, Zhang L, Gu L, Meng Q, Hou Y, Hu J. Health information interoperability and standard system—Practice of China. Health Policy and Technology 2014 Dec;3(4):248-252. [doi: 10.1016/j.hlpt.2014.10.003]
12. National Health Commission, People's Republic of China. National Population Health Informatization Development Plan for 13th Five-Year Plan. 2017 Feb 21. URL: http://www.nhc.gov.cn/guihuaxxs/s10741/201702/ef9ba6fbe2ef46a49c333de32275074f.shtml [accessed 2020-10-30]
13. Zhao F, Zhang P, Zhang Y, Ma Z. Time to lead the prevention and control of public health emergencies by informatics technologies in an information era. J Biosaf Biosecur 2020 Dec;2(2):102-104 [FREE Full text] [doi: 10.1016/j.jobb.2020.06.001] [Medline: 32954228]
14. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Apr 14;323(14):1341-1342. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
15. Ye Q, Zhou J, Wu H. Using Information Technology to Manage the COVID-19 Pandemic: Development of a Technical Framework Based on Practical Experience in China. JMIR Med Inform 2020 Jun 08;8(6):e19515 [FREE Full text] [doi: 10.2196/19515] [Medline: 32479411]
16. National Health Commission Statistical Information Center of the People's Republic of China. Survey Report on National Health Informatization. Beijing, China: People's Medical Publishing House; 2019.

17. Mäenpää T, Asikainen P, Gissler M, Siponen K, Maass M, Saranto K, et al. Outcomes assessment of the regional health information exchange: a five-year follow-up study. Methods Inf Med 2011 Feb;50(4):308-318. [doi: 10.3414/ME10-01-0048] [Medline: 21336419]
18. Mäenpää T, Suominen T, Asikainen P, Maass M, Rostila I. The outcomes of regional healthcare information systems in health care: a review of the research literature. Int J Med Inform 2009 Nov;78(11):757-771. [doi: 10.1016/j.ijmedinf.2009.07.001] [Medline: 19656719]
19. General Office of the National Health Commission, People's Republic of China. Notice of the General Office of the National Health Commission on Informatization Support for Normalized Epidemic Prevention and Control. 2020 Jun 28. URL: http://www.gov.cn/zhengce/zhengceku/2020-06/29/content_5522612.htm [accessed 2020-10-30]
20. Sittig DF, Singh H. COVID-19 and the Need for a National Health Information Technology Infrastructure. JAMA 2020 Jun 16;323(23):2373-2374. [doi: 10.1001/jama.2020.7239] [Medline: 32421178]
21. Iakovidis I. Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare record in Europe1Disclaimer: The view developed in this paper is that of the author and does not necessarily reflect the position of the European Commission. International Journal of Medical Informatics 1998 Oct;52(1-3):105-115. [doi: 10.1016/s1386-5056(98)00129-4]
22. Division of Primary Health, National Health Commission, People's Republic of China. Notice on Doing a Good Job in Basic Public Health Service Projects in 2019. 2019 Sep 4. URL: http://www.nhc.gov.cn/jws/s7881/201909/83012210b4564f26a163408599072379.shtml [accessed 2021-03-10]
23. Shen J, Li L. Considerations on rights related to electronic health records. Chinese Journal of Health Informatics and Management 2012 Apr;9(2):17-20.
24. Meng Q, Ye Q, Hu J. Regional Population Health Informatization Construction and Development. Beijing, China: People's Medical Publishing House; 2014.
25. National Health Commission, People's Republic of China. Answers to the Guiding Opinions on Accelerating the Construction of Population Health Informatization. 2013 Dec 9. URL: http://www.nhc.gov.cn/guihuaxxs/s10742/201312/2519dea9a4b14318a0736881116275ee.shtml [accessed 2020-10-30]
26. National Health Commission, People's Republic of China. Opinions on strengthening the construction of the national health information standardization system. 2020 Oct 10. URL: http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=4114443b613546148b275f191da4662b [accessed 2020-10-30]
27. Hu H, Qin P, Lei X. The Development Process and Outlook of National Health Informatization in China. J Med Inform 2019 Jul;40(7):2-6. [doi: 10.3969/j.issn.1673-6036.2019.07.001]
28. Wang K, Mao AY, Meng YL. Development history, current situation, problems and strategies of public health system construction in China. Chin J Public Health 2019 Jul;35(7):801-805.
29. Outline of National Health Informatization Development Plan 2003-2010. Health Supervision Center, National Health Commission, People's Republic of China. 2007 Nov 20. URL: https://www.jdzx.net.cn/article/4028818a14f4305e0114f43c1b0c0002/2009/3/4028818a163f0d8d01165ca36753005a.html [accessed 2020-10-30]
30. Cresswell K, Sheikh A, Franklin BD, Krasuska M, Nguyen HT, Hinder S, et al. Theoretical and methodological considerations in evaluating large-scale health information technology change programmes. BMC Health Serv Res 2020 May 27;20(1):477 [FREE Full text] [doi: 10.1186/s12913-020-05355-7] [Medline: 32460830]
31. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. Lancet Digit Health 2020 Apr;2(4):e166-e167 [FREE Full text] [doi: 10.1016/S2589-7500(20)30054-6] [Medline: 32289116]
32. Samra H, Li A, Soh B, Zain MA. Utilisation of hospital information systems for medical research in Saudi Arabia: A mixed-method exploration of the views of healthcare and IT professionals involved in hospital database management systems. Health Inf Manag 2020 May;49(2-3):117-126. [doi: 10.1177/1833358319847120] [Medline: 31046465]
33. Lee J. EMR management system for patient pulse data. J Med Syst 2012 Oct;36(5):3255-3260. [doi: 10.1007/s10916-011-9816-9] [Medline: 22223158]
34. Zhou G, Xu X, Hu J. Health Informatization: Development, Characteristics and Prospects. Chinese Journal of Health Informatics and Management 2019 Aug;16(4):384-394.
35. National Health Commission, People's Republic of China. Reply to Recommendation No. 8113 of the Second Session of the Thirteenth National People's Congress. 2020 Jul 22. URL: http://www.nhc.gov.cn/wjw/jiany/202007/082ceb95551b461fafdc85a3f725df5f.shtml [accessed 2020-10-30]
36. National Health Commission, People's Republic of China. China Health Statistical Yearbook. Beijing, China: China Union Medical College Press; 2019.
37. Hao XN, Ma CY, Lu ZY. The Effects and Problems on the Reform of Primary Health Informatization in China. Health Economics Research 2020 Jul;37(7):3-9. [doi: 10.14055/j.cnki.33-1056/f.2020.07.001]
38. China Hospital Information Management Association. Report on the Situation of Hospital Informatization in China in 2018-2019. 2019 Sep 1. URL: https://chima.org.cn/Html/News/Articles/4878.html [accessed 2020-10-30]
39. China Hospital Information Management Association. Research on the fund investment mechanism of regional universal health informatization (III). 2019 Oct 17. URL: https://www.chima.org.cn/Html/News/Articles/4845.html [accessed 2020-10-30]

40. Guo M, Wang B, Li J. International experience of government and social resource cooperation mechanisms in health informatization and its enlightenment. Chinese Health Resources 2017 Jan;20(1):78-82. [doi: 10.13688/j.cnki.chr.2017.16173]
41. Xia L, Qi X, Meng Y. Human resources for public health informatization in CDCs at provincial and municipal levels. Chin J Public Health 2013 Sep;29(9):1306-1310. [doi: 10.11847/zgggws2013-29-09-18]
42. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine During the COVID-19 Pandemic: Experiences From Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
43. Hasnain M, Pasha MF, Ghani I. Combined measures to control the COVID-19 pandemic in Wuhan, Hubei, China: A narrative review. J Biosaf Biosecur 2020 Dec;2(2):51-57 [FREE Full text] [doi: 10.1016/j.jobb.2020.10.001] [Medline: 33521592]
44. Sjödin H, Johansson AF, Brännström Å, Farooq Z, Kriit HK, Wilder-Smith A, et al. COVID-19 healthcare demand and mortality in Sweden in response to non-pharmaceutical mitigation and suppression scenarios. Int J Epidemiol 2020 Oct 01;49(5):1443-1453 [FREE Full text] [doi: 10.1093/ije/dyaa121] [Medline: 32954400]
45. Kho J, Gillespie N, Martin-Khan M. A systematic scoping review of change management practices used for telemedicine service implementations. BMC Health Serv Res 2020 Sep 01;20(1):815 [FREE Full text] [doi: 10.1186/s12913-020-05657-w] [Medline: 32873295]
46. Song X, Liu X, Wang C. The role of telemedicine during the COVID-19 epidemic in China-experience from Shandong province. Crit Care 2020 Apr 28;24(1):178 [FREE Full text] [doi: 10.1186/s13054-020-02884-9] [Medline: 32345359]
47. Liu L, Gu J, Shao F, Liang X, Yue L, Cheng Q, et al. Application and Preliminary Outcomes of Remote Diagnosis and Treatment During the COVID-19 Outbreak: Retrospective Cohort Study. JMIR Mhealth Uhealth 2020 Jul 03;8(7):e19417 [FREE Full text] [doi: 10.2196/19417] [Medline: 32568722]
48. National Health Commission, People's Republic of China. The regular press conference of the National Health Commission on September 9, 2020 introduced the typical cases of the application and development of national health informatization. 2020 Sep 9. URL: http://www.nhc.gov.cn/xwzb/webcontroller.do?titleSeq=11343&gecstype=1 [accessed 2020-10-30]
49. National Health Commission, People's Republic of China. Standardized maturity assessment of national medical and health information interconnection was carried out smoothly. 2019 Jul 8. URL: http://www.nhc.gov.cn/mohwsbwstjxxzx/s8553/201907/6c4292d3f95f4dc7b05334e867442d42.shtml [accessed 2020-10-30]
50. Ministry of Health, People's Republic of China. Notice of the General Office of the Ministry of Health on the Issuance of the Guidelines for the Construction of Regional Health Information Platform Based on Health Records (Trial). 2009 Jun 4. URL: http://www.nhc.gov.cn/wjw/gfxwj/201304/af448a0bee9e40648ca89d4d0b5f4ab9.shtml [accessed 2021-03-10]
51. Farokhzadian J, Khajouei R, Hasman A, Ahmadian L. Nurses' experiences and viewpoints about the benefits of adopting information technology in health care: a qualitative study in Iran. BMC Med Inform Decis Mak 2020 Sep 21;20(1):240 [FREE Full text] [doi: 10.1186/s12911-020-01260-5] [Medline: 32958042]
52. Han Y, Lie RK, Guo R. The Internet Hospital as a Telehealth Model in China: Systematic Search and Content Analysis. J Med Internet Res 2020 Jul 29;22(7):e17995 [FREE Full text] [doi: 10.2196/17995] [Medline: 32723721]
53. Jiang K, Feng J, Su Y. Prevention and Control Measures of Hospital Informatization from the Perspective of 'Fighting Epidemic'. Health Economics Research 2020 Nov;37(10):60-62. [doi: 10.14055/j.cnki.33-1056/f.2020.10.016]
54. Ina K, Furuta R, Kayukawa S, Iwasaki M, Sirokawa Y, Koga C, et al. Medical and dental cooperation using VPN with IPsec in the hospital without dentists. Annals of Oncology 2018 Oct;29:vii84. [doi: 10.1093/annonc/mdy375.083]
55. Ratwani RM, Reider J, Singh H. A Decade of Health Information Technology Usability Challenges and the Path Forward. JAMA 2019 Feb 26;321(8):743-744. [doi: 10.1001/jama.2019.0161] [Medline: 30715133]

Abbreviations

CT: computed tomography
DID: demographic information database
EHR: electronic health record
EHRD: electronic health record database
EMRD: electronic medical record database
HC: health commission
HIT: health information technology
ICU: intensive care unit
IT: information technology
NHC: National Health Commission
NHFPC: National Health and Family Planning Commission
SARS: severe acute respiratory syndrome
VPN: virtual private network

Edited by C Basch; submitted 24.01.21; peer-reviewed by G Fernandes, MA Bahrami; comments to author 05.03.21; revised version received 17.03.21; accepted 27.05.21; published 17.06.21.

Please cite as:
Huang M, Wang J, Nicholas S, Maitland E, Guo Z
Development, Status Quo, and Challenges to China’s Health Informatization During COVID-19: Evaluation and Recommendations
J Med Internet Res 2021;23(6):e27345
URL: https://www.jmir.org/2021/6/e27345
doi:10.2196/27345
PMID:34061761

Original Paper

eHealth Implementation Issues in Low-Resource Countries: Model, Survey, and Analysis of User Experience

Norman Archer[1*], PhD; Cynthia Lokker[2*], PhD; Maryam Ghasemaghaei[1*], PhD; Deborah DiLiberto[3*], PhD

[1]Information Systems, DeGroote School of Business, McMaster University, Hamilton, ON, Canada

[2]Health Research Methods, Evidence and Impact, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[3]Global Health, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Norman Archer, PhD
Information Systems
DeGroote School of Business
McMaster University
1280 Main St. West
Hamilton, ON, L8S 4M4
Canada
Phone: 1 905 525 9140 ext 23944
Email: archer@mcmaster.ca

Abstract

Background: The implementation of eHealth in low-resource countries (LRCs) is challenged by limited resources and infrastructure, lack of focus on eHealth agendas, ethical and legal considerations, lack of common system interoperability standards, unreliable power, and shortage of trained workers.

Objective: The aim of this study is to describe and study the current situation of eHealth implementation in a small number of LRCs from the perspectives of their professional eHealth users.

Methods: We developed a structural equation model that reflects the opinions of professional eHealth users who work on LRC health care front lines. We recruited country coordinators from 4 LRCs to help recruit survey participants: India, Egypt, Nigeria, and Kenya. Through a web-based survey that focused on barriers to eHealth implementation, we surveyed 114 participants. We analyzed the information using a structural equation model to determine the relationships among the constructs in the model, including the dependent variable, eHealth utilization.

Results: Although all the model constructs were important to participants, some constructs, such as user characteristics, perceived privacy, and perceived security, did not play a significant role in eHealth utilization. However, the constructs related to technology infrastructure tended to reduce the impact of concerns and uncertainties (path coefficient=−0.32; *P*=.001), which had a negative impact on eHealth utilization (path coefficient=−0.24; *P*=.01). Constructs that were positively related to eHealth utilization were implementation effectiveness (path coefficient=0.45; *P*<.001), the countries where participants worked (path coefficient=0.29; *P*=.004), and whether they worked for privately or publicly funded institutions (path coefficient=0.18; *P*<.001). As exploratory research, the model had a moderately good fit for eHealth utilization (adjusted R^2=0.42).

Conclusions: eHealth success factors can be categorized into 5 groups; our study focused on frontline eHealth workers' opinions concerning 2 of these groups: technology and its support infrastructure and user acceptance. We found significant disparities among the responses from different participant groups. Privately funded organizations tended to be further ahead with eHealth utilization than those that were publicly funded. Moreover, participant comments identified the need for more use of telemedicine in remote and rural regions in these countries. An understanding of these differences can help regions or countries that are lagging in the implementation and use of eHealth technologies. Our approach could also be applied to detailed studies of the other 3 categories of success factors: short- and long-term funding, organizational factors, and political or legislative aspects.

(J Med Internet Res 2021;23(6):e23715) doi:10.2196/23715

KEYWORDS

eHealth; low-resource countries; eHealth implementation effectiveness; end user survey; eHealth utilization

Introduction

Background

> *eHealth is the cost-effective and secure use of information communication technology in support of health and health-related fields, including health-care services, health surveillance, health literature, and health education, knowledge and research [1].*

Properly implemented eHealth has the potential to scale up the delivery of health care to people in low-resource countries (LRCs) [2]. Quality of health care has been found to contribute the most to the success of these eHealth interventions, whereas cost contributes the most to intervention failures [3]. The challenges faced when implementing eHealth vary among countries. For eHealth solutions to succeed in LRCs, an organized approach must be used to address these challenges.

A 2016 World Health Organization report [1] indicates significant progress in eHealth planning:

> *More than half of WHO Member States now have an eHealth strategy, and 90% of eHealth strategies reference the objectives of UHC [Universal Health Coverage] or its key elements. It is becoming mainstream for countries to have policies for managing information. When well-articulated, eHealth strategies should enable the interoperability needed to support people-centred health services for everyone, and the move from disease silos to resilient health systems which can deliver UHC.*

Although the World Health Organization claims that more than half of their member states have an eHealth strategy, actual implementation of their strategies is not always followed. This is clear from published studies and concerns expressed by many researchers [3-7]. For example, Kiberu et al [4] suggest that although many sub-Saharan African countries are evaluating eHealth as a means of improving health care accessibility, several are engaged in the proof-of-concept stage of unsustainable pilot projects. There are no national guidelines in many LRCs for the secure management of individual digital health information and services, placing personal data at risk. Implementation issues with standards and interoperability can create barriers to the use of eHealth and its spread across regions or nations to support the full realization of potential health system benefits. The potential for eHealth to reduce health care costs and enable access to better quality health care is limited, often due to inadequate funding, inadequate infrastructure causing power blackouts, poor internet connectivity, and an unskilled eHealth workforce. However, regardless of having to work in such difficult circumstances, progress is being made in implementing eHealth in many LRCs. For example, a few of the many research papers that have been published by LRC researchers include critical issues such as eHealth being used to combat infant mortality in rural and remote regions of India [8] and Nigeria [9], open-source electronic health record systems that support interoperable links among them have been installed in Kenya [10], and telecom policies developed to encourage optimal digital network implementation to support eHealth in Egypt [11].

Objective

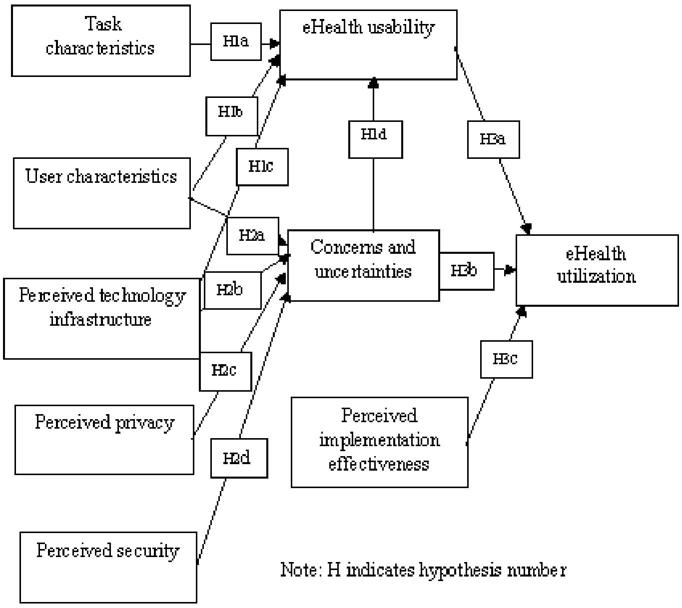
The objective of this study is to create and validate a model of the factors that influence the successful implementation of eHealth in LRCs, based on eHealth challenges identified in a survey of LRC eHealth end users. A recent categorization [12], modified by Ahmed et al [13], has synthesized eHealth success factors into 5 categories: (1) technology and its support infrastructure, (2) user acceptance, (3) short- and long-term funding, (4) organizational factors, and (5) political or legislative aspects. Our study focuses on factors 1 and 2, which are likely to be of interest and intimately familiar to the end user participants we surveyed.

Methods

Construct Development for the eHealth Implementation Issues Model

Figure 1 shows the model used in this study. The following discussion describes the details of the model's development and related hypotheses (shown in small rectangles in the diagram), including the background of the model constructs. The study based on this model was implemented through a web-based questionnaire detailed in Multimedia Appendix 1.

Figure 1. Structural equation model of eHealth implementation issues in low-resource countries.



Note: H indicates hypothesis number

Task Characteristics

Tasks are the activities that support the organization. Information systems facilitate completion of the organization’s tasks. Compatibility with work processes or work styles and task difficulty are often used to characterize tasks [14]. Research has shown [15] that if a system is more aligned with the needs of the users, there is a greater chance of system success. A fit-variability model [16] (related to the task-technology fit model by Goodhue [15]) showed that different stakeholders may perceive eHealth viability and fit of the same eHealth services very differently and that there can be discrepancies between organizational viability and individual fit of specific eHealth services. Relevant information from the study by Goodhue and Thompson [15] was used to derive the task characteristics construct for the questionnaire (Multimedia Appendix 1). This leads to:

- Hypothesis 1a: Task characteristics will positively affect eHealth usability.

User Characteristics

“User characteristics are attitudes, perceptions, and demographics that are specific to the individual users of the information system” [14]. A survey of 465 medical professionals in northwest Nigerian hospitals [7] showed that the majority of the participants had a good level of literacy for implementing and working with new eHealth systems. They found statistically significant positive correlations between intention to use eHealth and attitude toward eHealth, perceived usefulness toward eHealth, information technology experience and eHealth, and technical infrastructure for eHealth. The user characteristics construct for the questionnaire (Multimedia Appendix 1) was adapted from the study by Zayyad and Toycan [7]. This leads to:

- Hypothesis 1b: User characteristics will positively affect eHealth usability.
- Hypothesis 2a: User characteristics will negatively affect concerns and uncertainties.

Perceived Technology Infrastructure

Information communications and technology is a key component of any eHealth program. For an eHealth program to succeed, technology must be available to support a database that is always directly and easily accessible to practitioners for health record input, retrieval, analysis, and sharing within the patient’s circle of care. The infrastructure must include stable and reliable digital communications network hardware and software to support health record access and sharing through point-of-care devices used by health care providers. Highly reliable electrical power must also be available at all times to drive the components of the infrastructure [4,6,7,17]. The relevant perceived technology infrastructure construct for the questionnaire (Multimedia Appendix 1) was adapted from the study by Zayyad and Toycan [7] as follows:

- Hypothesis 1c: Perceived technology infrastructure will positively affect eHealth usability.
- Hypothesis 2b: Perceived technology infrastructure will negatively affect concerns and uncertainties.

Perceived Privacy

A study of eHealth in Bangladesh found that privacy concerns by patients [18] did not have a significant impact on provider concerns about eHealth use. An explanation is that poor eHealth privacy and security considerations resulted in eHealth systems being judged by patients as failures [19]. This is the opposite of the findings from users of eHealth (clinicians) who may tend to believe that privacy and security are less important when an eHealth system provides superior health care. The perceived privacy construct for the questionnaire (Multimedia Appendix 1) was derived from the study by Wilkowska and Ziefle [20], as follows:

- Hypothesis 2c: Perceived privacy will positively affect concerns and uncertainties.

Perceived Security

Security and privacy issues are related in eHealth systems; it is not possible to manage privacy without a secure system. A survey of professionals working in Nigerian hospitals addressed the professionals' intentions to use eHealth [7]. A question related to security concerns showed a nonsignificant correlation of this issue with the intention to use eHealth. The perceived security construct of the questionnaire (Multimedia Appendix 1) was derived from the study by Wilkowska and Ziefle [20] as follows:

- Hypothesis 2d: Perceived security will positively affect concerns and uncertainties.

eHealth Usability

Usability represents an important yet often overlooked factor impacting the implementation and meaningful use of eHealth systems. Without usable systems, doctors, medical technicians, nurses, administrative staff, and other users would have great difficulty in realizing the potential benefits of eHealth systems. The usability of technical systems has been studied in the information systems field, beginning with the landmark work by Davis [21]. This is a key measure of an eHealth system and directly reflects how users may react positively to its use. In general, if the system is built to perform specific user tasks, its usability will be greater. The eHealth usability construct for the questionnaire (Multimedia Appendix 1) was adapted from the study by Davis [21]. The foregoing leads to the following:

- Hypothesis 3a: eHealth usability will positively affect eHealth utilization.

Concerns and Uncertainties About eHealth

A review of empirical research classifying eHealth implementations as successes or failures [3] found that quality of health care was most often mentioned as contributing to the success of eHealth interventions. This review found that costs were most frequently mentioned as contributing to failure, although workflow issues were also mentioned in most of the articles reviewed. Workflow issues could lead to disagreement among the affected clinicians, increasing uncertainties, and the potential for failure of the eHealth system. The concerns and uncertainties about eHealth construct of the questionnaire (Multimedia Appendix 1) adapted ideas expressed by Aranda-Jan et al [22] in a systematic review of what does not work in African eHealth projects, leading to the following hypotheses:

- Hypothesis 1d: Concerns and uncertainties about eHealth systems will negatively affect eHealth usability.
- Hypothesis 3b: Concerns and uncertainties about eHealth systems will negatively affect eHealth utilization.

Perceived Implementation Effectiveness

Underlying factors affect health care professionals' decisions to implement eHealth technology applications in LRCs [7]. These include the perceived usefulness, belief, willingness, and attitude of health care professionals. Our study implicitly reflects these factors in terms of survey feedback from users who have chosen to implement eHealth in their workplaces. The technological capability of eHealth systems is one of the key factors that influence the successful implementation of a technology [17]. Technological success factors include functional and nonfunctional requirements, interoperability, and user interface design. The long-term sustainability of a system depends on the economic, social, and organizational sustainability in which the technology is embedded. The perceived eHealth implementation effectiveness construct of the questionnaire (Multimedia Appendix 1) was developed from ideas expressed by Rezai-Rad et al [23] and is stated as:

- Hypothesis 3c: Perceived implementation effectiveness will positively affect eHealth utilization.

eHealth System Utilization

Utilization of an eHealth system is a measure of how popular the system is with the users and if it will be sustainable and worth the operating cost in the long run. This was measured in our study by eHealth system utilization, a one-indicator formative construct Q10 (Multimedia Appendix 1) that lists possible eHealth utilization levels by the participant's organization.

Table 1 summarizes the reference sources mentioned above that were used to create the eHealth implementation model.

Table 1. Summary of sources for eHealth implementation model constructs.

Title	Construct	Type	Study
Perceived task characteristics	Validated	Reflective	Goodhue and Thompson [15]
Individual characteristics	New	Reflective	Zayyad and Toycan [7]
Perceived technology infrastructure	New	Reflective	Zayyad and Toycan [7]
Perceived eHealth privacy	Validated	Reflective	Wilkowska and Ziefle [20]
Perceived eHealth security	Validated	Reflective	Wilkowska and Ziefle [20]
eHealth usability	Validated	Reflective	Davis [21]
Concerns and uncertainties about eHealth	New	Reflective	Aranda-Jan et al [22]
eHealth implementation effectiveness	New	Reflective	Rezai-Rad et al [23]
eHealth utilization	New	Formative (1-indicator variable)	N/A[a]

[a] N/A: not applicable (as this construct is developed in this study).

Implementation

Overview

This study was approved by the McMaster University Research Ethics Board. In addition to its own approval process, the board contacted eHealth authorities in each of the 4 LRCs to ensure that the research proposal was acceptable. The study proceeded in 2 phases: I and II. An individual country coordinator with experience in eHealth implementations was recruited for each of the 4 countries involved in the study. They received nominal reimbursement for managing the recruitment of eHealth expert consultants for phase I and survey participants for phase II in their respective countries. eHealth consultants in the phase 1 study also received a nominal sum. Participants were not paid for completing the survey (phase II). In the 2 countries where 35 or more participants were recruited, a random draw prize was awarded to 1 participant in each country.

Phase I

On the basis of a detailed review of the relevant eHealth literature, we developed a draft questionnaire. Our study focused on eHealth users and support staff in the 4 representative LRCs: Kenya (East Africa), Nigeria (West Africa), India (South Asia), and Egypt (North Africa). Consultations with the experts on the phase I questionnaire informed the design of the final questionnaire and highlighted the importance of the factors identified from the literature. A key finding was that the model was too broad, including a number of strategic concerns that individual participants, as clinicians or other end users, were less likely to be directly concerned with. These strategic concerns (3 of the 5 categories [12] referenced earlier in this paper: short- and long-term funding, organizational factors, and political or legislative aspects) made the questionnaire too long for busy eHealth users to be willing to complete. Therefore, we reduced the scope of the issues covered to the first 2 categories (technology and its support infrastructure and user acceptance). The final questionnaire was based on the more limited model shown in Figure 1. Each reflective construct in the model included at least three indicator variables, which were presented on a 7-point Likert scale. The resulting questionnaire is provided in Multimedia Appendix 1, and the information, consent, and invitation to participate messages to the survey participants are provided in Multimedia Appendix 2. The web-based version, developed using *Qualtrics* software[24], took about 15 minutes for the participants to complete.

Phase II

A convenience sampling survey of eHealth users in the 4 LRCs was arranged by the relevant country coordinators, who recruited suitable participants. Participants were from public and private institutions in rural and urban areas and had varying levels of eHealth experience. Details of the survey process are provided in Multimedia Appendix 3. From 177 invitations to participate, 114 (64.4% overall response rate) valid responses were completed from India (39/114, 34.2%), Egypt (52/114, 45.6%), Kenya (11/114, 9.6%), and Nigeria (12/114, 10.5%). Statistical data were analyzed using partial least squares with *Smart PLS3* software [25].

Results

Participant Demographics

Participant demographics are detailed in Table 2, including country comparisons. In the table, the *Total* column represents the number of participants from each country who completed the survey successfully. All data are presented as absolute values and percentages. Percentages in the total column on the righthand side sum to 100% for the categories presented in each of the 5 table divisions that were also used, along with country, as control variables (occupation, employer, eHealth experience, urban or rural experience, and employment experience).

Table 2. Participant demographics.

Characteristics	Country, n (%)				Total (N=114), n (%)
	India (n=39)	Egypt (n=52)	Kenya (n=11)	Nigeria (n=12)	
Occupation					
Physicians	8 (20.5)	20 (38.4)	1 (9.1)	1 (8.3)	81 (71.1)
Nurses	1 (2.5)	1 (1.9)	1 (9.1)	0 (0)	8 (7)
Allied health personnel	4 (10.2)	3 (5.8)	1 (9.1)	1 (8.3)	25 (21.9)
Employer					
Work in privately funded health care	10 (25.4)	10 (19.2)	1 (9.1)	1 (8.3)	61 (53.5)
Work in publicly funded health care	3 (7.7)	14 (26.9)	1 (9.1)	1 (8.3)	53 (46.5)
eHealth experience					
No previous experience with eHealth	4 (10.2)	7 (13.4)	0 (0)	0 (0)	26 (22.8)
2 or more years of experience with eHealth	6 (15.3)	19 (36.5)	2 (18.2)	1 (8.3)	88 (77.2)
Urban versus rural experience					
eHealth experience only in urban settings	1 (2.5)	10 (19.2)	1 (9.1)	0 (0)	28 (24.6)
eHealth experience only in rural settings	0 (0)	0 (0)	0 (0)	0 (0)	2 (1.8)
eHealth experience in both rural and urban settings	12 (31.6)	14 (26.9)	1 (9.1)	1 (8.3)	84 (73.6)
Employment experience					
Predominant eHealth experience in clinics	0 (0)	6 (11.5)	1 (9.1)	1 (8.3)	26 (22.7)
Predominant eHealth experience in education	1 (2.5)	8 (15.3)	0 (0)	1 (8.3)	23 (20.6)
Predominant eHealth experience in technology support	3 (7.7)	3 (5.8)	0 (0)	0 (0)	16 (13.7)
Predominant eHealth experience in training	0 (0)	5 (9.6)	0 (0)	0 (0)	12 (10.5)
Predominant eHealth experience in monitoring and evaluation	1 (2.5)	7 (13.4)	1 (9.1)	0 (0)	20 (18)
Predominant eHealth experience in administration	1 (2.5)	2 (3.8)	1 (9.1)	0 (0)	10 (8.6)
Predominant eHealth experience in planning	0 (0)	3 (5.8)	0 (0)	0 (0)	7 (5.9)

eHealth Implementation Model Results

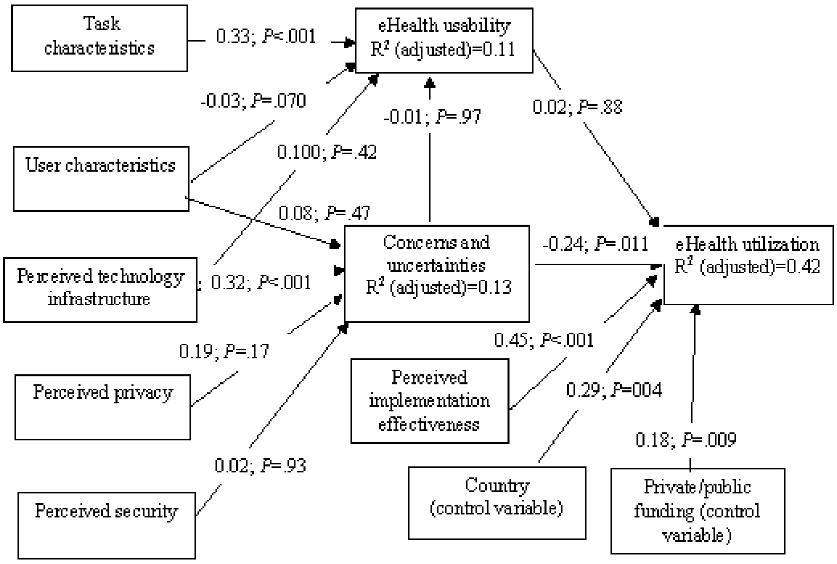
Figure 2 shows the results calculated from the structural equation model, which was run with bootstrapping using 1000 subsamples. The calculated path coefficients for the proposed hypotheses, shown in Figure 2, are listed in Table 3. Hypotheses H1a, H2b, H3b, and H3c were supported, whereas the remaining hypotheses (H1b, H1c, H1d, H2a, H2c, H2d, and H3a) were not supported (all with *P*>.05).

Control variables (for the demographic categories in Table 2) were also run against the model, and those with significant results are included in Figure 2. These are (1) participant's country and (2) employer private or public funded. The results for these control variables are shown in the lower part of Table 3.

The composite reliabilities and average variance extracted (AVE) for the reflective constructs are shown in Table 4. The composite reliability of a construct measures the reliability of the indicator variables included in the construct. All the composite reliabilities were above the accepted lower limit of 0.70 [26]. The AVE results measure the fit of the internal structure of the model. AVE is slightly below the accepted lower limit of 0.50 for *concerns and uncertainties*, but all the other values meet the lower limit within the rounding error; therefore, with this exception, the model has convergent validity. The heterotrait ratio of correlations [27], shown in Table 5, assesses the discriminant validity in the model. The resulting maximum value of 0.79 is below the 0.85 threshold, so discriminant validity is established. The adjusted R^2 values for usability, concerns and uncertainties, and eHealth utilization are listed in Table 6.

Participants' responses to the *eHealth utilization* construct were analyzed according to the extent to which the participants indicated that eHealth was used in their organization. The results, stratified by country, are shown as absolute values and percentages in Table 7.

Figure 2. Model results for eHealth implementation issues.



Note: Path data are coefficients with *P* values

Table 3. Calculated path coefficients and significance.

Relationship	Path coefficient	*P* value
Hypothesis		
H1a: Task characteristics→[a]eHealth usability	0.33	<.001
H1b: User characteristics→eHealth usability	−0.03	.87
H1c: Perceived technology infrastructure→eHealth usability	0.10	.42
H1d: Concerns and uncertainties about eHealth→eHealth usability	0.01	.97
H2a: User characteristics→concerns and uncertainties	−0.08	.47
H2b: Perceived technology infrastructure→concerns and uncertainties	−0.32	<.001
H2c: Perceived privacy→concerns and uncertainties	0.20	.17
H2d: Perceived security→concerns and uncertainties	0.02	.93
H3a: Usability→eHealth utilization	0.02	.88
H3b: Concerns and uncertainties→eHealth utilization	−0.24	.01
H3c: Perceived implementation effectiveness→eHealth utilization	0.45	<.001
Significant control variables		
Country of participant→eHealth utilization	0.29	.004
Private or public funding→eHealth utilization	0.18	.009

[a] Arrows represent the directional relationships of the coefficients.

Table 4. Composite reliabilities and average variance extracted for reflective constructs.

Construct	Composite reliability	AVE[a]
Concerns and uncertainties	0.74	0.42
Perceived implementation effectiveness	0.89	0.73
Perceived privacy	0.88	0.71
Perceived security	0.72	0.49
Perceived technology infrastructure	0.86	0.68
Perceived usability	0.85	0.59
Task characteristics	0.78	0.48
User characteristics	0.89	0.72

[a] AVE: average variance extracted.

Table 5. Discriminant analysis via heterotrait-monotrait ratio of correlations.

Constructs	Concerns and uncertainty	Perceived implementation effective	Perceived privacy	Perceived security	Perceived technology infrastructure	Task characteristics	Usability	User characteristics
Perceived implementation effectiveness	0.36	—[a]	—	—	—	—	—	—
Perceived privacy	0.36	0.16	—	—	—	—	—	—
Perceived security	0.17	0.14	0.79	—	—	—	—	—
Perceived technology infrastructure	0.46	0.54	0.04	0.10	—	—	—	—
Task characteristics	0.33	0.77	0.25	0.23	0.57	—	—	—
Usability	0.19	0.32	0.27	0.24	0.25	0.48	—	—
User characteristics	0.25	0.53	0.14	0.10	0.44	0.48	0.15	—
eHealth utilization	0.35	0.53	0.02	0.18	0.44	0.55	0.33	0.24

[a] Not applicable.

Table 6. Adjusted R^2 from model calculations.

Latent variable	Adjusted R^2
Usability	0.12
Concerns and uncertainties	0.13
eHealth utilization	0.42

Table 7. Participants' responses to the question "indicate to what extent eHealth is used in your organization."

Extent of eHealth use in my organization	Total (N=114), n (%)	Egypt (n=52), n (%)	India (n=39), n (%)	Nigeria (n=12), n (%)	Kenya (n=11), n (%)
Never	3 (2.6)	3 (5.8)	0 (0)	0 (0)	0 (0)
To a very small extent	22 (19.3)	12 (23.1)	9 (23.1)	0 (0)	1 (9.1)
To a small extent	24 (21.1)	12 (23.1)	4 (10.2)	5 (41.6)	3 (27.3)
To a moderate extent	35 (30.7)	14 (26.9)	15 (38.5)	4 (33.3)	2 (18.2)
To a fairly great extent	17 (14.9)	8 (15.4)	6 (15.4)	2 (16.7)	1 (9.1)
To a great extent	6 (5.3)	3 (5.8)	2 (5.1)	0 (0)	1 (9.1)
To a very great extent	7 (6.1)	0 (0)	3 (7.7)	1 (8.3)	3 (27.3)

Discussion

Principal Findings

The focus of our study was on providers and support staff with applied experience in eHealth and the extent of use they perceived of eHealth in their organizations. The structural equation model we developed resulted in an estimate of the extent of eHealth utilization in their organizations (adjusted R^2=0.42). This result is in the range of a moderately good result [28] for an exploratory study.

In the model, only task characteristics contributed significantly to usability (Figure 2). When the model was run without control variables in place, it gave an adjusted R^2 of 0.310 for eHealth utilization, with the path coefficient usability→eHealth utilization significant at 0.18 (*P*=.01). Other path coefficients changed very little, but this one became nonsignificant at a value of 0.015 (*P*=.88) when the 2 significant control variables (country and public or private employment) were included in the final model. This result tells us that there was variability among the opinions of the participants about eHealth utilization that depended upon their home country and their employer (private or public).

eHealth Utilization

eHealth utilization in participant organizations was dealt with in question 10 of Multimedia Appendix 1. Its raw data results were analyzed and are presented in Table 7. These results show a similar distribution of engagement level of eHealth in health care organizations in each of the 4 countries. The median use of eHealth by the LRC participant organizations was *to a moderate extent*, as indicated by calculations from the raw data in response to the question in Multimedia Appendix 1*the extent of eHealth use in your organization*. This suggests a generally favorable reaction to the applicability of eHealth to health care organization work. Although the sample sizes of the Nigerian and Kenyan responses were too small for statistical comparisons, there was little difference in their average results from the Egyptian and Indian responses.

Privacy and Security

Privacy and its related supporting functionality, that is, security, were not found to be significant to the model construct eHealth concerns and uncertainties and, thus, to eHealth utilization, which is in agreement with similar studies [7,18]. Although this may be the case for eHealth users such as our participants, patients themselves in other LRC studies were found to be concerned about privacy violations through secondary or unauthorized access [19,29]. However, the lack of significance in our model did not mean that privacy and security were unimportant to the participants. From the raw data for question 4 in Multimedia Appendix 1, the overall result was a median value of 7 (you strongly agree) and a mean value of 6.3 (you agree) for the 3 positive statements in the questionnaire about the relevance of privacy, and similar results for the 3 positive statements in the questionnaire about the relevance of security (Multimedia Appendix 1). These are favorable results that did not have a significant impact on the model results because they did not seem to be of concern to most participants in relation to eHealth utilization.

Relevance to Previous Literature

The recent history of eHealth implementation and experimentation in many LRCs has resulted in research and publication of many relevant results, including those that touched on sustainable implementation of eHealth in these countries [3,7,17,18,30,31]. In addition, Mauco et al [32] developed and validated an eHealth readiness assessment framework for developing countries. It includes a comprehensive set of readiness factors, including organizational, technological or infrastructural, government, societal, health care provider, engagement, core, and public- or patient-related. Another study by Ahmed et al [13], adapted from the study by Broens et al [12], synthesized 5 mobile health and telehealth (generalizable to eHealth) success factors, which we referenced in the introduction to this paper. Our research focused on 2 of these success factors (technology and its support infrastructure and user acceptance) to portray their effects on user perceptions.

The implementation of eHealth systems in LRCs differs from past activities with the more mature systems in developed nations. For example, many LRCs have been implementing pilot eHealth systems, some of which have been successful whereas others have not. Some of these implementations have ignored long-term effects, such as nonstandard systems that do not interoperate with other existing or planned systems [4]. Mistakes of this nature were also made when eHealth systems were initially being used in developed nations, and it is important to avoid making the same expensive mistakes in LRCs.

Other eHealth Implementation Measures

Whether hospitals implement experimental or full eHealth applications, their operations are a source of data and user opinions that could be harvested to deduce predictions and possible causes of success or failure in future installations. The problem is that almost all related research on LRCs has been based on single installations or systematic reviews [3,7,22,23,32], and it is difficult to generalize from these to validate a theoretical framework. We note that there are organizations that publish hospital rankings in different regions and countries, including most, if not all, LRCs (eg, *Ranking Web of Hospitals* [33]). These rankings do not specifically include eHealth considerations. More specific to hospital implementations of eHealth is the Electronic Medical Record Adoption Model (EMRAM; HIMSS Analytics) [34]. EMRAM is an 8-level maturity model, beginning at level 0 (no eHealth facilities) and improving to level 7 (virtually complete implementation of eHealth, including electronic medical records, external digital links, privacy, security, disaster recovery, data analytics, and data governance). Forward-looking hospital managers aspire to move upward on this scale. Although few hospitals outside the United States have reached levels 6 or 7, many hospitals in the United States have reached levels 6 or 7. Although the EMRAM approach has been used primarily in developed nations, its advantage is that it encourages hospitals to modernize their facilities through eHealth implementation in a carefully managed manner. It also helps the hospital management to look for advice from other hospitals to help them move more effectively to a higher level of eHealth implementation. The need for this more organized approach is similar in LRCs, except that LRC hospitals are often starting at an introductory, low level of eHealth use, where initial eHealth system adoption requires extensive changes in hospital operations and employee training.

Implications for Future Design

Our research takes the first step in the use of local eHealth experience by combining the opinions of actual eHealth end users from several LRCs. Regardless of the digital design of eHealth systems with which participants worked, the results of this study can be generalized to other proposed installations. Our findings could ultimately influence the design of eHealth systems, apps, and interfaces.

We also believe that we have demonstrated in a small way how to improve the general theory of eHealth implementations in LRCs by assessing a simultaneous combination of opinions of end users about 2 of the main eHealth success factors [13]. By redesigning the model; extending the survey scope; and expanding the participant audience to users, planners, developers, decision makers, and politicians, an evaluation of all 5 success factors [13] in multiple eHealth installations could be undertaken. This would help LRC planners, aided by a modified maturity model approach, to develop an appreciation of the impact of the various factors [35] that may differ among LRCs or among different installations in the same LRC, through the expressed opinions of participants. This would also help to direct strategic investments in eHealth more effectively.

Strengths and Limitations

Our eHealth research in LRCs was the first time this approach was used to gather users' perceptions of how eHealth utilization differed among the countries involved in the study. Acknowledging some limitations in our survey design, it is clear that studies of this nature with revised survey formats could be undertaken on an expanded scale, involving participants with a wider range of eHealth backgrounds. The ultimate gain would be a wider development and understanding of an approach similar to maturity modeling to help the hospital management move ahead with eHealth implementation in an organized and optimal manner.

We received many interesting and useful comments from participants that we were unable to analyze and include here because of space limitations.

Limitations included our use of convenience sampling to identify participants, which was not fully representative of each country's health care workforce. The participant response rate was also considerably less than statistically desirable, and we could have achieved a higher response rate if we had been able to pay participants a nominal fee.

Acknowledgments

This research was supported by a joint grant from the McMaster University Faculty of Health Sciences, the DeGroote School of Business, and the McMaster University Research Excellence Fund. The authors wish to thank them for their generous support. We also relied on our support network members to oversee the gathering of data from eHealth users in the four countries in our study. The support network included Adekunle Ajiboye, MEEI, MSc, CEO AAJIMATICS, Hamilton, Canada; Victor Gbenro, BMBCh, PgD, MSc, Dep. Registrar, Medical & Dental Council Nigeria, Abuja, Nigeria; and the four country research coordinators Chythra Rao, MD, Kasturba Medical College, Manipal, India; Ahmed Shaarawy, MD, MSc, MBBcH, Clinical Informatics Director, Welcare Hospital, Giza, Egypt; Wycliffe Ayieko, PgD, University of Nairobi, Kenya; and Nkem Momah, MBBS, MPH, Deputy Director, eGovernment Development & Regulatory Dept, National Information & Technology Development Agency, Abuja, Nigeria. Finally, we want to thank the participants in the 4 countries who completed the questionnaires for this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire.
[DOCX File , 42 KB - jmir_v23i6e23715_app1.docx]

Multimedia Appendix 2
Information, consent, and invitation to participate in the online survey.
[DOCX File , 30 KB - jmir_v23i6e23715_app2.docx]

Multimedia Appendix 3
Details of the survey process.
[DOCX File , 22 KB - jmir_v23i6e23715_app3.docx]

References

1. Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the center for health market innovations. Bull World Health Org 2012 May 1;90(5):332-340. [doi: 10.2471/blt.11.099820]
2. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
3. Global Diffusion of EHealth: Making Universal Health Coverage Achievable: Report of the Third Global Survey on EHealth. Geneva, Switzerland: World Health Organization; 2016.
4. Kiberu VM, Mars M, Scott RE. Barriers and opportunities to implementation of sustainable e-Health programmes in Uganda: A literature review. Afr J Prim Health Care Fam Med 2017 May 29;9(1):e1-e10 [FREE Full text] [doi: 10.4102/phcfm.v9i1.1277] [Medline: 28582996]
5. Mauco KL, Scott RE, Mars M. Critical analysis of e-health readiness assessment frameworks: suitability for application in developing countries. J Telemed Telecare 2016 Dec 23;24(2):110-117. [Medline: PMID: 28008790]
6. Omotosho A, Ayegba P, Emuoyibofarhe J, Meinel C. Current state of ICT in healthcare delivery in developing countries. Int J Onl Eng 2019 May 14;15(08):91. [doi: 10.3991/ijoe.v15i08.10294]
7. Zayyad M, Toycan M. Factors affecting sustainable adoption of e-health technology in developing countries: an exploratory survey of Nigerian hospitals from the perspective of healthcare professionals. PeerJ 2018;6:e4436 [FREE Full text] [doi: 10.7717/peerj.4436] [Medline: 29507830]
8. Venkatesh V, Rai A, Sykes TA, Aljafari R. Combating infant mortality in rural India: evidence from a field study of ehealth kiosk implementations. MISQ 2016 Feb 2;40(2):353-380. [doi: 10.25300/misq/2016/40.2.04]
9. Obasola OI, Mabawonku IM. Mothers' perception of maternal and child health information disseminated via different modes of ICT in Nigeria. Health Info Libr J 2018 Dec 24;35(4):309-318 [FREE Full text] [doi: 10.1111/hir.12235] [Medline: 30251307]
10. Muinga N, Magare S, Monda J, Kamau O, Houston S, Fraser H, et al. Implementing an open source electronic health record system in Kenyan health care facilities: case study. JMIR Med Inform 2018 Apr 18;6(2):e22 [FREE Full text] [doi: 10.2196/medinform.8403] [Medline: 29669709]
11. Badran MF. eHealth in Egypt: the demand-side perspective of implementing electronic health records. Telecommunications Policy 2019 Jul;43(6):576-594. [doi: 10.1016/j.telpol.2019.01.003]
12. Broens TH, Huis in't Veld RM, Vollenbroek-Hutten MM, Hermens HJ, van Halteren AT, Nieuwenhuis LJ. Determinants of successful telemedicine implementations: a literature study. J Telemed Telecare 2007;13(6):303-309. [doi: 10.1258/135763307781644951] [Medline: 17785027]
13. Ag Ahmed MA, Gagnon M, Hamelin-Brabant L, Mbemba GI, Alami H. A mixed methods systematic review of success factors of mhealth and telehealth for maternal health in Sub-Saharan Africa. Mhealth 2017;3:22 [FREE Full text] [doi: 10.21037/mhealth.2017.05.04] [Medline: 28736731]
14. Petter S, DeLone W, McLean ER. Information systems success: the quest for the independent variables. J Manag Inf Syst 2014 Dec 8;29(4):7-62. [doi: 10.2753/mis0742-1222290401]
15. Goodhue D, Thompson R. Task tchnology fit and individual performance. MIS Q 1995;27(3):213-236.
16. Mettler T. Anticipating mismatches of HIT investments: developing a viability-fit model for e-health services. Int J Med Inform 2016 Jan;85(1):104-115. [doi: 10.1016/j.ijmedinf.2015.10.002] [Medline: 26526279]
17. Fanta GB, Pretorius L. A conceptual framework for sustainable ehealth implementation in resource-constrained settings. South African J Indust Eng 2018 Nov;29(3):132-147. [doi: 10.7166/29-3-2055]
18. Hoque MR, Bao Y, Sorwar G. Investigating factors influencing the adoption of e-Health in developing countries: a patient's perspective. Inform Health Soc Care 2017 Jan 11;42(1):1-17. [doi: 10.3109/17538157.2015.1075541] [Medline: 26865037]
19. Adu EK, Todorova N, Mills A. Do Individuals in Developing Countries Care about Personal Health Information Privacy? An Empirical Investigation. AIS E-Library. 2020. URL: https://aisel.aisnet.org/confirm2019/16 [accessed 2021-06-02]
20. Wilkowska W, Ziefle M. Privacy and data security in E-health: requirements from the user's perspective. Health Informatics J 2012 Sep;18(3):191-201 [FREE Full text] [doi: 10.1177/1460458212442933] [Medline: 23011814]
21. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
22. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
23. Rezai-Rad M, Vaezi R, Nattagh F. E-health readiness assessment framework in iran. Iran J Public Health 2012;41(10):43-51 [FREE Full text] [Medline: 23304661]
24. Meet the Software That Powers More Than 1 Billion Surveys Every Year. Qualtrics. 2020. URL: https://www.qualtrics.com/uk/core-xm/survey-software/ [accessed 2021-05-29]
25. SmartPLS. SmartPLS is here. Smart PLS. 2020. URL: https://www.smartpls.com/ [accessed 2021-05-21]
26. Hair JF, Sarstedt M, Ringle CM, Mena JA. An assessment of the use of partial least squares structural equation modeling in marketing research. J Acad Mark Sci 2011 Jun 7;40(3):414-433. [doi: 10.1007/s11747-011-0261-6]

27. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
28. Wong K. Estimating complex models using higher order construct modeling (HCM). In: Mastering Partial Least Squares Structural Equation Modeling (PLS-SEM) With Smartpls in 38 Hours. Bloomington, Indiana: Universe; 2019:87-106.
29. Pankomera R, van Greunen D. Privacy and Security Issues for a Patient-centric Approach in Public Healthcare in a Resource Constrained Setting. In: Africa Week Conference. 2016 Presented at: AWC'16; May 6-9, 2016; Durban, South Africa. [doi: 10.1109/istafrica.2016.7530664]
30. Isabalija SR, Mbarika V, Kituyi GM. A framework for sustainable implementation of e-medicine in transitioning countries. Int J Telemed Appl 2013;2013:615617 [FREE Full text] [doi: 10.1155/2013/615617] [Medline: 24454353]
31. Siribaddana P, Hewapathirana R, Jayatilleke AU, Sahay S, Dissanayake VH. Strengthening health systems through informatics capacity development among doctors in low-resource contexts: the Sri Lankan experience. South East Asia J Public Health 2019 Sep;8(2):87-94 [FREE Full text] [doi: 10.4103/2224-3151.264852] [Medline: 31441443]
32. Mauco KL, Scott RE, Mars M. Validation of an e-health readiness assessment framework for developing countries. BMC Health Serv Res 2020 Jun 23;20(1):575 [FREE Full text] [doi: 10.1186/s12913-020-05448-3] [Medline: 32576174]
33. Ranking Web of World Hospitals. Webometrics. 2020. URL: https://hospitals.webometrics.info/en/Methodology [accessed 2021-05-21]
34. Electronic Medical Record Adoption Model (EMRAM). HIMSS Analytics. 2020. URL: https://www.himssanalytics.org/emram [accessed 2021-05-21]
35. Luna D, Almerares A, Mayan JC, González Bernaldo de Quirós F, Otero C. Health Informatics in Developing Countries: Going beyond Pilot Practices to Sustainable Implementations: A Review of the Current Challenges. Healthc Inform Res 2014 Jan;20(1):3-10 [FREE Full text] [doi: 10.4258/hir.2014.20.1.3] [Medline: 24627813]

Abbreviations

AVE: average variance extracted
EMRAM: Electronic Medical Record Adoption Model
LRC: low-resource country

Edited by R Kukafka, G Eysenbach; submitted 20.08.20; peer-reviewed by H Hochheiser, K Yin; comments to author 26.09.20; revised version received 01.11.20; accepted 15.05.21; published 18.06.21.

Please cite as:
Archer N, Lokker C, Ghasemaghaei M, DiLiberto D
eHealth Implementation Issues in Low-Resource Countries: Model, Survey, and Analysis of User Experience
J Med Internet Res 2021;23(6):e23715
URL: https://www.jmir.org/2021/6/e23715
doi:10.2196/23715
PMID:34142967

Original Paper

The Impact of Technology-Enabled Care Coordination in a Complex Mental Health System: A Local System Dynamics Model

Frank Iorfino[1], PhD; Jo-An Occhipinti[1], PhD; Adam Skinner[1], PhD; Tracey Davenport[1], eMBA; Shelley Rowe[1], PhD; Ante Prodan[2], PhD; Julie Sturgess[3], BNurs, MN; Ian B Hickie[1], MD

[1]Brain and Mind Centre, University of Sydney, Sydney, Australia
[2]Translational Health Research Institute, Western Sydney University, Sydney, Australia
[3]North Coast Primary Health Network, Coffs Harbour, Australia

Corresponding Author:
Frank Iorfino, PhD
Brain and Mind Centre
University of Sydney
94 Mallet Street, Camperdown
Sydney, 2050
Australia
Phone: 61 2 9351 0584
Email: frank.iorfino@sydney.edu.au

Abstract

Background: Prior to the COVID-19 pandemic, major shortcomings in the way mental health care systems were organized were impairing the delivery of effective care. The mental health impacts of the pandemic, the recession, and the resulting social dislocation will depend on the extent to which care systems will become overwhelmed and on the strategic investments made across the system to effectively respond.

Objective: This study aimed to explore the impact of strengthening the mental health system through technology-enabled care coordination on mental health and suicide outcomes.

Methods: A system dynamics model for the regional population catchment of North Coast New South Wales, Australia, was developed that incorporated defined pathways from social determinants of mental health to psychological distress, mental health care, and suicidal behavior. The model reproduced historic time series data across a range of outcomes and was used to evaluate the relative impact of a set of scenarios on attempted suicide (ie, self-harm hospitalizations), suicide deaths, mental health–related emergency department (ED) presentations, and psychological distress over the period from 2021 to 2030. These scenarios include (1) business as usual, (2) increase in service capacity growth rate by 20%, (3) standard telehealth, and (4) technology-enabled care coordination. Each scenario was tested using both pre– and post–COVID-19 social and economic conditions.

Results: Technology-enabled care coordination was forecast to deliver a reduction in self-harm hospitalizations and suicide deaths by 6.71% (95% interval 5.63%-7.87%), mental health–related ED presentations by 10.33% (95% interval 8.58%-12.19%), and the prevalence of high psychological distress by 1.76 percentage points (95% interval 1.35-2.32 percentage points). Scenario testing demonstrated that increasing service capacity growth rate by 20% or standard telehealth had substantially lower impacts. This pattern of results was replicated under post–COVID-19 conditions with technology-enabled care coordination being the only tested scenario, which was forecast to reduce the negative impact of the pandemic on mental health and suicide.

Conclusions: The use of technology-enabled care coordination is likely to improve mental health and suicide outcomes. The substantially lower effectiveness of targeting individual components of the mental health system (ie, increasing service capacity growth rate by 20% or standard telehealth) reiterates that strengthening the whole system has the greatest impact on patient outcomes. Investments into more of the same types of programs and services alone will not be enough to improve outcomes; instead, new models of care and the digital infrastructure to support them and their integration are needed.

(J Med Internet Res 2021;23(6):e25331) doi:10.2196/25331

KEYWORDS

medical informatics; internet; care coordination; complex systems; simulation; health systems; policy; mental health

Introduction

Mental illness is a major cause of disability and early mortality globally [1-3]. The early onset, chronicity, and complexity of mental illness means that the human, social, and economic costs are enormous [4]. Yet, mental health systems often struggle to provide adequate care that effectively intervenes to mitigate the short- and long-term negative outcomes [5].

For some, access to mental health care has improved [6-9]; however, most health systems lack effective coordination between service silos, which impacts on the delivery of holistic, timely, and quality mental health care [10,11]. Service fragmentation, delayed care, mental health treatment isolated from other physical and social needs, complicated service pathways, and inefficient resource allocation are persistent features of an underperforming service system [5,12]. This is particularly problematic for people with more complex needs who tend to report higher rates of disengagement from education or employment [7-9], suicidality [13,14], and comorbidity [15]. These cases are the norm among those who frequently engage with health systems and typically require multidisciplinary team-based care approaches [16]. These approaches embrace collaborative care models, which recognize that effective care coordination between service providers, including intensive assessment, personalized treatment plans, targeted referrals, clinical information systems use, and outcome monitoring can improve treatment engagement, satisfaction with care, and mental health outcomes [17-19].

The integrated use of digital technologies offers significant potential to enable effective coordination of mental health care [20-23]. The accessibility, scalability, and standardization mean that technology is well-placed to play a major role in the digitization of health care and can be leveraged to deliver quality mental health care across settings, especially those that are remote or rural and may be underresourced when compared to urban centers [24]. Their use in mental health systems has already demonstrated utility to improve access to care and communication between service providers and consumers [25]. This may be particularly important for people with more complex needs, such as housing and employment support, whereby mental health treatment may be an additional burden in terms of time, effort, and finance. Technology offers an opportunity to alleviate some of this burden through greater flexibility in terms of access to effective care, greater efficiencies through the sharing of information between providers, and improving engagement with care [26].

Never has this capability been more vital than amid a pandemic and recession whose effects are disrupting nearly every aspect of life: familial, educational, vocational, health, and social structures. This disruption is threatening population mental health and well-being and is likely to generate service demand of unprecedented magnitude for many years [27,28]. Some governments are responding by instituting measures to reduce economic and social hardship, investing in mental health programs and services, and improving access via virtual mental health services (ie, basic telehealth). These investments represent a move in the right direction; however, without addressing the fundamental models of care, the rapid adoption of telehealth poses the risk that it may digitize the problems that already exist in existing models of care across the mental health system [29,30].

Systems modeling and simulation is a low-risk method of exploring likely impacts of counterfactual scenarios. This study uses an existing local system dynamics model developed for regional mental health services planning to identify the impact of using digital technologies to facilitate care coordination on mental health outcomes and health system burden (ie, technology-enabled care coordination). Here, we aimed to compare technology-enabled care coordination to three other likely intervention scenarios: (1) business as usual, (2) an increase in the growth rate of existing service capacity (ie, increasing throughput of people into existing systems), and (3) digital technologies used to extend existing services online (eg, via videoconferencing) without changing the underlying model of care (ie, standard telehealth). We aimed to model the impacts of each scenario within both a *typical* context and a public mental health crisis resulting from the pandemic and economic recession.

Methods

Context

The North Coast NSW (New South Wales) Primary Health Network supports a population of 502,524, as of 2016 [31], distributed over a geographic area of approximately 35,570 square kilometers and takes in both coastal and inland rural communities. The region is more socioeconomically disadvantaged compared to the state and national averages, with higher rates of unemployment, domestic and family violence, and homelessness [32,33].

Model Overview

The system dynamics model was developed using a participatory modeling approach that involved over 50 local stakeholders, including representatives from health and social policy agencies, local government, nongovernmental organizations, primary care providers, emergency services, research institutions, community groups, and, importantly, people with lived experience. The process employed a broad systems perspective drawing on the deep tacit knowledge of the diverse perspectives of these system actors. Input from stakeholders was provided through a series of workshops, meetings, surveys, and local system mapping activities between July and December 2019. This process involved iteratively working on the model structure and assumptions; regular face validity checks by a diverse group of academic, clinical, policy, program planning, emergency services, and lived-experience stakeholders were undertaken to ensure accurate model representation, conceptualization, and outputs. A more detailed description of the model development process, structure, outputs, and calibration can be found in the primary paper [34].

In summary, the core model structure included the following dynamically interacting components:

1. A population component, capturing changes over time in the size and structure of the population resulting from births,

net migration, and mortality across the following age groups: 0-14 years, 15-24 years, 25-64 years, and 65+ years old.

2. A psychological distress component that captures changes in the rates and severity of psychological distress in the population (ie, states of low psychological distress, with a Kessler 10 [K10] score of 10-15; moderate psychological distress, with a K10 score of 16-21; and high to very high psychological distress, with a K10 score of 22-50).
3. A series of components capturing pathways within and between key social determinants of mental health and suicidal behavior, namely, early life exposures, substance abuse, domestic violence, homelessness, and unemployment.
4. A mental health service system component that models the movement of psychologically distressed people through one of several possible service pathways, potentially involving general practitioners; psychiatrists; allied mental health professionals, including psychologists and mental health nurses; emergency department (ED) and psychiatric inpatient care; community- and hospital-based outpatient care; and online services, and one that captures changes over time in service demand and capacity.
5. A suicidal behavior component that captures self-harm hospitalizations and suicide deaths. Figure S1 in Multimedia Appendix 1 details the key social determinants identified by the participatory process and their hypothesized impact on mental health and the other model outcomes.

Model construction and analysis were performed using Stella Architect, version 1.9.4 (isee systems inc). The model was validated by (1) testing whether the model could replicate historic data across a range of key indicators, namely, time series of psychological distress, psychiatric hospitalizations, mental health–related ED presentations, self-harm hospitalizations, and suicide deaths, and (2) ensuring face validity among diverse system actors in the model structure and performance. The AdViSHE (Assessment of the Validation Status of Health-Economic decision models) checklist was also used to assess the validation status of this model and its outputs [35].

Model Outputs

For this study, model outputs included both mental health outcomes and service usage for the total population. Mental health outcomes and health system burden include total cumulative numbers of self-harm hospitalizations, which are a proxy for suicide attempts; suicide deaths; prevalence of high to very high psychological distress; and mental health–related ED hospitalizations. All outputs were calculated every 0.4375 days (ie, one-sixteenth of a week) starting from January 1, 2011; these permitted comparisons of model outputs with historic data from 2011 to 2017 for validation. Forecasts of the impacts of intervention scenarios described below are simulated from the time of implementation in 2021 to the end of 2029 [34].

Intervention Scenarios

Four distinct intervention scenarios were tested. The first scenario is *business as usual*, whereby there is no change to the existing services system and the rate of growth in existing services is maintained. The second scenario is *an increase in service capacity growth by 20%*, whereby the yearly growth of existing service capacity is increased. These services include those provided by general practitioners; psychiatrists; allied mental health professionals, including psychologists and mental health nurses; and community-based mental health services. The third scenario is *standard telehealth*, whereby digital technologies are used to extend existing services online (eg, via videoconferencing) without changing the underlying model of care. The fundamental assumption here is that technologies are being used to remotely provide existing models of care delivered by individual providers rather than being used to improve the coordination of care between service providers: a coordinated, multidisciplinary, team-based approach to care. This means that there is no change to referrals between services nor to the per-service recovery rate, since the only element that has changed is the mode of delivery. The fourth scenario is *technology-enabled care coordination*, which involves the use of online technology to facilitate delivery of multidisciplinary team-based care, in which medical and allied health professionals consider all relevant treatment options and collaboratively develop an individual treatment and care plan for each patient. Online technology enables enhanced coordination of care and facilitates communication between medical and allied health professionals, since each health professional involved in the care of a patient has access to the same information about that patient's treatment history (see parameter details in Textbox 1 [17,25,34,36,37]).

Each scenario was operationalized by varying key parameter estimates in the model, as outlined in Table 1. Each of these intervention scenarios was tested under *pre–COVID-19 conditions*, which are based on original trends in socioeconomic circumstances (ie, all policies and programs currently in place remain unchanged and service capacity continues to increase at current rates). Each intervention scenario was also tested under *post–COVID-19 conditions*, which reflect a new baseline for socioeconomic circumstances to compare and contrast the qualitative and quantitative performance of this intervention when faced with new realities (eg, socioeconomic shocks). The interventions considered here are not intended to be a complete evaluation of strategies to mitigate the effects of the COVID-19 pandemic, but rather we aimed to determine the robustness of each scenario under these changed conditions. Table 2 details how the pre– and post–COVID-19 initial conditions were implemented in the model.

Textbox 1. Parameters determining the direct effects of technology-enabled care coordination [34].

Maximum rate per service:

This refers to the maximum proportion of mental health services provided that involve technology-enabled care coordination. This proportion will depend on the number of medical and allied mental health professionals adopting technology-enabled care coordination, as well as the number of patients consenting to the use of these technologies in the management of their care (ie, uptake among service providers and patients). The default value (0.7) assumes that when fully implemented, technology-enabled care coordination will be provided in 70% of mental health services completed.

Effect on recovery rate:

This refers to the multiplicative effect of technology-enabled care coordination on the per-service recovery rate (ie, the probability that a patient's level of psychological distress will decrease after receiving treatment). The default estimate (1.177) implies that technology-enabled care coordination increases the per-service probability of a reduction in psychological distress by 17.7% [17,25].

Effect on referrals to specialized care:

This refers to the multiplicative effect of technology-enabled care coordination on general practitioners' rates of referral to specialized psychiatric care (ie, psychiatrists and allied mental health services). The default value (1.266) implies that technology-enabled care coordination increases the per-consultation probability that a general practitioner will refer a patient with high or very high psychological distress to specialized psychiatric care by 26.6% [36]. Note that technology-enabled care coordination is assumed to have no effect on the referral rate for patients with moderate psychological distress.

Effect on disengagement:

This refers to the multiplicative effect of technology-enabled care coordination on rates of disengagement from mental health services, including waiting for services and dissatisfaction with services received. The default estimate (0.520) implies that technology-enabled care coordination reduces rates of disengagement by 48.0% [25,36].

Effect on referrals to alcohol and other drugs services:

This refers to the multiplicative effect of technology-enabled care coordination on the rate of referral of patients with a substance abuse disorder to alcohol and drug treatment services. The default value (1.1) assumes an increase in the rate of referral of 10% (ie, patients with a substance abuse disorder receiving technology-enabled care coordination are 10% more likely to be referred to alcohol and drug treatment services than patients with a substance abuse disorder receiving usual care).

Effect on substance use relapse rate:

This refers to the multiplicative effect of coordinated treatment of co-occurring substance abuse and mental disorders on the substance use relapse rate (ie, the probability that a patient treated for a substance use disorder will relapse when treatment is completed). The default value (0.869) implies that coordinated substance abuse and psychiatric treatment reduces the rate of substance use relapse by 13.1% (ie, compared to substance abuse treatment alone) [37].

Effect on employment initiation:

This refers to the multiplicative effect of technology-enabled care coordination on the rate at which unemployed patients gain employment, through referral to employment services. The default value (1.1) assumes an increase in the employment initiation rate of 10% (ie, unemployed patients receiving technology-enabled care coordination are 10% more likely to gain employment than unemployed patients receiving usual care).

Effect on exiting homelessness rate:

This refers to the multiplicative effect of technology-enabled care coordination on the rate at which homeless patients secure housing, through referral to homelessness services. The default value (1.1) assumes a 10% increase in the rate that patients exit homelessness, equal to the inverse of the duration of homelessness (ie, homeless patients receiving technology-enabled care coordination are 10% more likely to secure housing than homeless patients receiving usual care).

Effect on psychiatric service capacity:

This refers to the multiplicative effect of technology-enabled care coordination on the total capacity of specialized psychiatric services (ie, the maximum number of services that can be provided by psychiatrists and allied mental health providers per year). The default value (1.1) assumes an increase in service capacity of 10%.

Effect on referrals to online services:

This refers to the multiplicative effect of technology-enabled care coordination on the rate of referral of patients with moderate psychological distress to online services. The default value (1.1) assumes an increase in the rate of referral of 10% (ie, patients with moderate psychological distress receiving technology-enabled care coordination are 10% more likely to be referred to online services than patients with moderate psychological distress receiving usual care).

Table 1. Parameter values for each intervention scenario.

Parameter[a]	Parameter values for each scenario			
	Scenario 1[b]	Scenario 2[c]	Scenario 3[d]	Scenario 4[e]
Direct effects on technology interventions				
Maximum rate per service	0	0	0.70	0.70
Effect on recovery rate	1.00	1.00	1.09	1.18
Effect on referrals to specialized care	1.00	1.00	10.00	1.27
Effect on disengagement	1.00	1.00	0.76	0.52
Effect on alcohol and other drugs services referral rate	1.00	1.00	1.00	1.10
Effect on substance abuse relapse	1.00	1.00	0.93	0.87
Effect on employment	1.00	1.00	1.00	1.10
Effect on exiting homelessness rate	1.00	1.00	1.00	1.10
Effect on service capacity	1.00	1.00	1.10	1.10
Effect on referrals to online services	1.00	1.00	1.00	1.10
Direct effects on service capacity				
General practice service capacity increase per year	125.85	151.02	125.85	125.85
Psychiatrist and allied service capacity increase per year	216.31	259.57	216.31	216.31
CMHC[f] service capacity increase per year	0	75.14	0	0

[a] See Textbox 1 for more details about each parameter.

[b] Scenario 1: business as usual.

[c] Scenario 2: increase in service capacity growth rate by 20%.

[d] Scenario 3: standard telehealth.

[e] Scenario 4: technology-enabled care coordination.

[f] CMHC: community mental health capacity.

Table 2. Comparison of parameters used to model pre– and post–COVID-19 initial conditions.

Parameter	Parameter values for each condition	
	Pre–COVID-19	Post–COVID-19
Youth job loss rate ratio	1.00	5.00
Unemployment total	1.00	15.00
Unemployment effect decay rate	0	0.05
Sense of community index (ie, social connectedness)	9.61	8.24
Years to reach sense of community index	0	1.00
Duration (years) of social disconnection	0	2.00

Statistical Analysis

Sensitivity analyses were performed to assess the impact of uncertainty in parameter estimates of the direct effects of each intervention scenario on the simulation results. Latin hypercube sampling was used to draw 100 sets of values for selected model parameters determining the direct effects of the interventions on suicidal behavior in young people, from a uniform joint distribution spanning ±20% of the default values. Differences in each of the model outputs (eg, cumulative numbers of self-harm hospitalizations, suicide deaths, and mental health–related ED presentations) between the baseline and intervention scenarios were calculated for each set of parameter values and were summarized using simple descriptive statistics. All intervals reported in this paper are derived from the distributions of model outputs calculated in the sensitivity analyses; they provide a measure of the impact of uncertainty in the intervention effect estimates but should not be interpreted as confidence intervals.

Results

Overview

Under the baseline scenario (ie, business as usual pre–COVID-19), approximately 12,274 self-harm hospitalizations, 953 suicide deaths, and 81,263 mental health–related ED presentations were forecast for the period from 2021 to 2030. By 2030, the prevalence of high psychological distress was expected to be on a trajectory of decline and was set to reduce by 2 percentage points (ie, from 17.60% in 2021 to 15.60% in 2030). Figure 1 shows the relative impact of each scenario on mental health and suicide outcomes. The implementation of technology-enabled care coordination (scenario 4) had the largest impact on outcomes for the total population. For the total population, self-harm hospitalizations and suicide deaths were reduced by 6.71% (95% interval 5.63%-7.87%), mental health–related ED presentations were reduced by 10.33% (95% interval 8.58%-12.19%), and the prevalence of high psychological distress in 2030 was reduced by 1.76 percentage points (95% interval 1.35-2.32 percentage points). This would result in the prevention of 844 self-harm hospitalizations, 66 suicides, and 8448 mental health–related hospitalizations over the forecast period and a decline in the prevalence of high psychological distress to 14% in 2030. The rate of uptake of technology-enabled care coordination has a major impact on the effects of this intervention (Figure 2). The impact on all outcomes is greater as the rate of uptake across the mental health system increases from 20% to 50% and then to 80%.

Figure 1. Mental health outcomes for each intervention scenario. ED: emergency department.

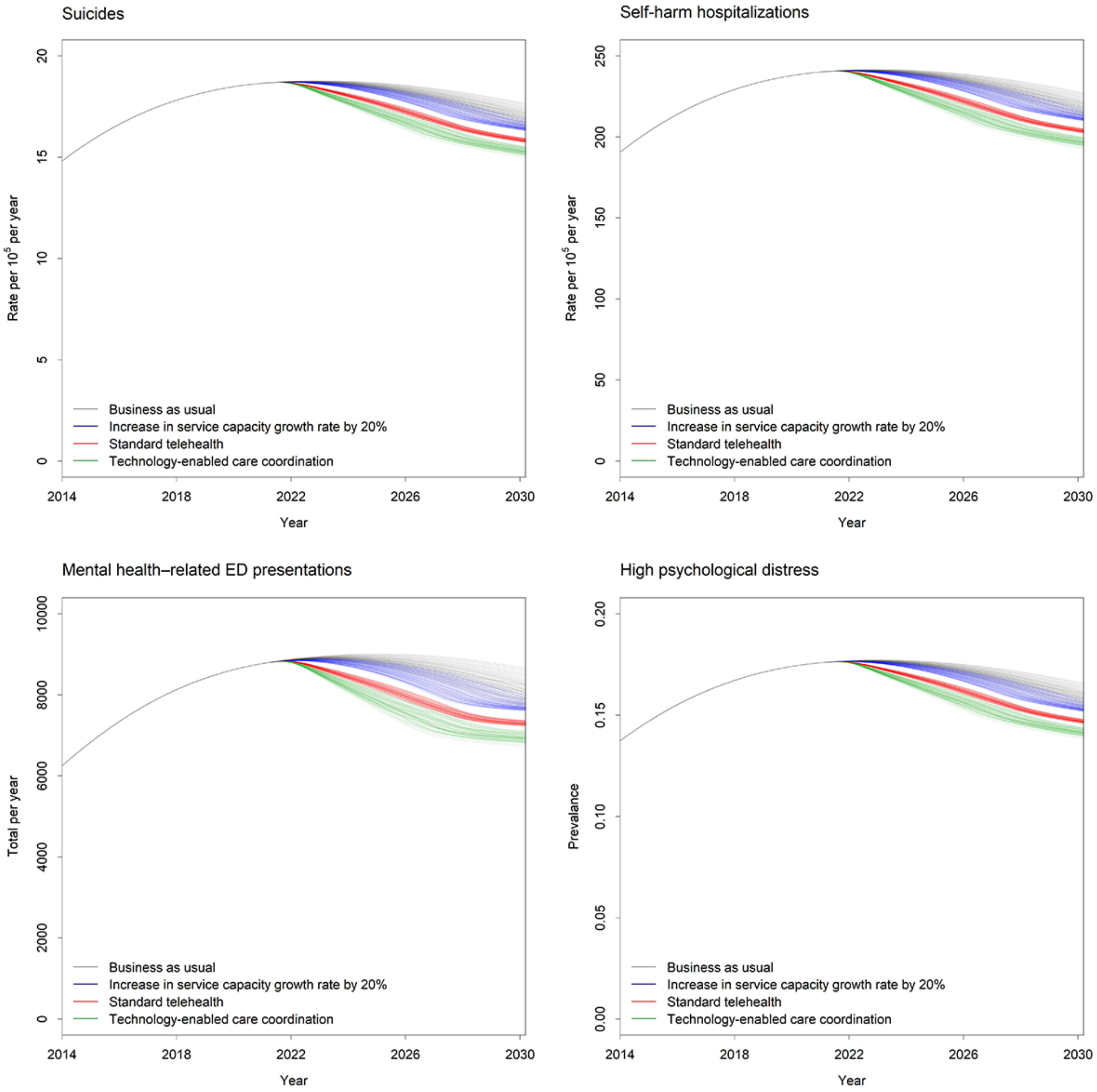
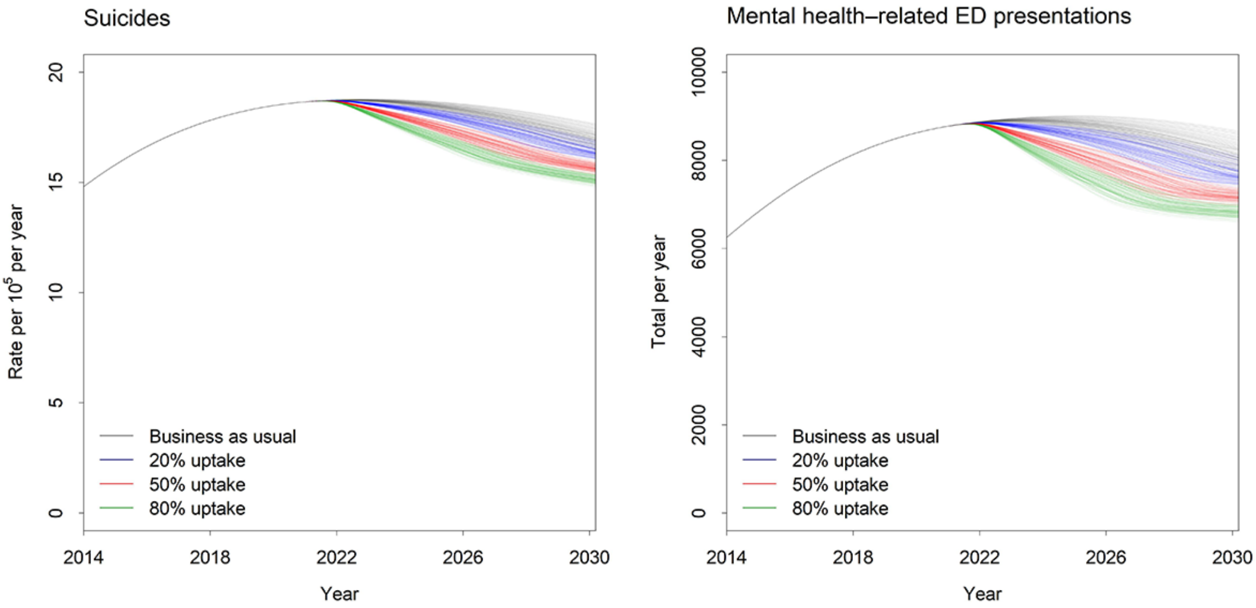


Figure 2. The impact of the rate of uptake on mental health outcomes. ED: emergency department.



Increasing the service capacity growth rate by 20% (scenario 2) only had a minor impact on both populations; for the total population, it was forecast to reduce self-harm hospitalizations and suicide deaths by 1.64% (95% interval –0.20% to 3.51%), mental health–related ED presentations by 2.76% (95% interval –0.41% to 5.95%), and the prevalence of high psychological distress in 2030 by 0.57 percentage points (95% interval –0.08 to 1.29 percentage points). This would result in the prevention of 274 self-harm hospitalizations, 21 suicides, and 3023 mental health–related hospitalizations. The prevalence of high psychological distress was projected to decline to 15% in 2030.

When digital technologies were used for standard telehealth (scenario 3), the impact on outcomes was better than scenario 2 but lower than scenario 4, reducing self-harm hospitalizations and suicide deaths by 3.50% (95% interval 1.58%-5.03%), mental health–related ED presentations by 5.48% (95% interval 2.24%-5.43%), and the prevalence of high psychological distress in 2030 by 0.99 percentage points (95% interval 0.42-1.67 percentage points). Over the forecast period, this would result in the prevention of 418 self-harm hospitalizations, 32 suicides, 4244 mental health–related hospitalizations, and a decline in the prevalence of high psychological distress to 14.8% in 2030.

Post–COVID-19 Scenario Modeling

The impact of the public mental health crisis was estimated to have detrimental effects on all outcomes, with 14,973 self-harm hospitalizations (22% increase from baseline figures), 1176 suicide deaths (23% increase), and 98,591 mental health–related ED presentations (21% increase). By 2030, the prevalence of high psychological distress was forecast to be 18.3%, a 2.7–percentage point increase from the baseline figures forecast for 2030. During this public mental health crisis, technology-enabled care coordination (scenario 4) was forecast to have a similar impact on outcomes, with a 6.43% (95% interval 5.29%-7.46%) reduction in self-harm hospitalizations, a 6.40% (95% interval 5.26%-7.42%) reduction in suicide deaths, a 10.01% (95% interval 8.06%-11.71%) reduction in mental health–related ED presentations, and a reduction of 2.52 percentage points (95% interval 1.82-3.08 percentage points) in high psychological distress in 2030. The pattern of results for each scenario was consistent under these changed conditions; technology-enabled care coordination represents an important and effective part of the strategy to mitigate the social and economic impacts of the COVID-19 pandemic, with outcomes reduced back down to the initial *business as usual* estimates from pre–COVID-19.

Discussion

Principal Findings

These findings suggest that the use of technology-enabled care coordination is likely to result in better mental health outcomes and reduce health system burden at a population level. When compared to increasing the service capacity growth rate across a variety of settings by 20% and standard telehealth, technology-enabled care coordination led to greater reductions in suicide deaths, the total number of self-harm hospitalizations or mental health–related ED hospitalizations, and the prevalence of high psychological distress in the population. This emphasizes that strengthening how the whole mental health system functions together will have a greater impact on outcomes than simply improving the capacity across individual components of the mental health system. Investments into more of the same types of programs and services alone will not be enough to improve the outcomes for the whole system; instead, new models of care and the infrastructure to support them and their integration need to accompany these investments.

Increasing the service capacity growth rate does have a minor impact; however, since it primarily acts to increase throughput into the current mental health system, it does not address existing challenges of service fragmentation. Many structural barriers and misaligned incentives remain across the system, which contribute to the health system burden and inefficient allocation of resources that often result in poor outcomes for individuals

[38-40]. Addressing these systemic issues will most likely involve ensuring that the mental health system has the appropriate infrastructure in place to not only meet the demand for services, but provide timely care that appropriately targets the diverse needs of people presenting for care.

This work supports calls for increased use of digital technologies in mental health care [41]; however, it also suggests the importance of the way these technologies are employed. The modeling shows that when digital technologies are used for standard telehealth practices by extending existing services online (eg, via videoconferencing), without changing the underlying model of care, then the impact is lower. This type of scenario reflects what we might expect to see when telehealth is more widely implemented to deliver existing services, yet little effort is made to utilize these technologies in ways that promote multidimensional team-based care and maximize the benefits that technologies provide. As such, while telehealth stands to improve access, reach, and throughput to the mental health system and is a critical improvement in health service delivery, it does not necessarily ensure that people receive the appropriate type of mental health and social care that improves their chances of recovery. For this, a transformation in the models of care provided by services within the mental health system are required to achieve the full potential benefits of digital technologies.

Research regarding the effectiveness of digital technologies for mental health is growing, yet many of these innovations focus on the use of technologies within closed systems of care, often in isolation from other parts of the mental health system [30,42]. This neglects the way people typically need to access multiple parts of the system to receive effective care. This study illustrates that the benefits of technology-enabled care coordination for the whole system continues to increase as the proportion of services using this intervention increases. Unfortunately, realizing this type of widespread usage will require overcoming specific implementation barriers that have plagued most attempts to implement new technologies into existing health systems [22,43]. Common barriers include technology design, variation in the level of integration into existing service pathways and clinical protocols, process dynamics, contextual factors (eg, local leadership and organizational support), and other factors (eg, resourcing and training) [44-46]. Addressing these barriers requires a whole system approach that challenges the traditional and often rigid health systems to ensure that these tools are developed and integrated with services in a way that truly transforms clinical practice for the whole system.

This work should be considered in light of some limitations. The use of multiple secondary data sets introduces potential measurement bias for the estimates used to parameterize the model. Strategies were used to reduce the impact of such biases, including the triangulation of multiple data sources, parameter estimation via constrained optimization, and local verification to identify plausible estimates. The impacts of simulated scenarios are not necessarily generalizable to other regions, due to the specificity of the population, demographic, behavioral, social, economic, and mental health service dynamics that drive outcomes in a particular modeled region. Yet, for regions contextually similar to the modeled region, many of the model insights are likely to be relevant and provide a compelling case for exploring the likely impacts of similar technologies elsewhere. Further, the distinctions between the assumptions of scenarios 3 and 4 are likely to provide more generalizable insights about what components of digital transformation are projected to have large impacts in health systems. Future work should also focus on validating these model outputs over time by embedding them in local monitoring and evaluation of the implemented technology. In addition, comparisons of the modeled impacts of the technology applied to regions with similar or different social, economic, and mental health system contexts will provide additional broader insights. Finally, this study focuses on population effects of scenarios whereby the effects of interventions are generalized at the population level and cannot account for individual differences in the way that digital technologies may be rolled out or implemented at a local level. To address these specific issues, individual-level approaches, such as agent-based modeling, may be more appropriate to determine the effect of digital technologies on different agents within a mental health system (eg, people, clinicians, and services) and the impact this has on mental health and suicide outcomes.

Conclusions

Systems modeling and simulation of the likely impact of technology-enabled care coordination in ordinary and extraordinary times has highlighted its significant potential in improving population mental health and suicide outcomes. This work also provides important evidence to support a push for major investment to scale up the implementation of digital technologies that support new models of care facilitating care coordination.

Acknowledgments

The model was developed in partnership with the North Coast Primary Health Network and was intended as a contribution to the North Coast Collective, which is a regional collaboration initially between three organizations—North Coast Primary Health Network, Northern NSW Local Health District, and Mid-North Coast Local Health District—and is growing to include partners and a range of community stakeholders, including those outside of the health sector and people with lived experience of mental ill-health and suicidal behavior. Critically, this regional focus seeks to deliver on regionally agreed outcomes, optimizing the intervention and investment portfolio to achieve the greatest gain—value is defined by the quadruple aim of health care—in the most efficient way. This work was made possible by generous contributions of time, local knowledge, data, and content area expertise by the North Coast Collective modeling group, which includes the following members: Fay Adamson, Dr Graeme Browne, Dr Ashlea Broomfield, Luke Buckley, Warren Burgess, Dale Cara, Vigeeta Chauban, Christopher Clark, Dr Trish Collie, Janet Cormick, Katrina Dart, Swati Devmurari, Samara Finlayson, Guy Flaherty, Robyn George, Dr Adrian Gilliland, Dr Nicola Holmes, Rebecca Hope, David Horseman, Danica Hudson, Glen James, Darren Kershaw, Nicholas Kosseris, Deborah Kuhn, Dr Toney Lower, Dr Corrine Maynard, Tracey Mills, Lyn Moran, Niall Mulligan, David Noble, Stanley Nebo, Stuart Nunan, Judy O'Mara, Margarete Anna Ogonowski, Anne-Maree Parry, Alan Pretty, Leavina Reid, Rachel Richardson, Deidre Robinson, Vicki Rose, Shelly Rowe, Sara Shaughnessy, Dr Peter Silberberg, Janny Sinclair, Liz Stirling, Julie Sturgess, Trent Taylor, Claudine Tule, Dr Richard Tranter, Darryl Walker, Tracey Wild, Leigh Williamson, Diana Wilson, and Jami-Lee Winnacott.

Conflicts of Interest

IBH was an inaugural commissioner on Australia's National Mental Health Commission (2012-2018). He is the Co-Director, Health and Policy at the Brain and Mind Centre (BMC), University of Sydney. The BMC operates an early-intervention youth services program at Camperdown under contract to Headspace. He is the Chief Scientific Advisor to, and a 5% equity shareholder in, InnoWell Pty Ltd. InnoWell was formed by the University of Sydney (45% equity) and PwC (Australia; 45% equity) to deliver the Aus $30 million (~US $23,212,500) Australian Government–funded Project Synergy (2017-2020; a 3-year program for the transformation of mental health services) and to lead the transformation of mental health services internationally through the use of innovative technologies. JAO is Head of Systems Modelling, Simulation & Data Science and Co-Director of the Mental Wealth Initiative at the University of Sydney's BMC, and is Managing Director of Computer Simulation & Advanced Research Technologies (CSART). TD is the Director of Research & Insights at the Australian Digital Health Agency.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 8212 KB - jmir_v23i6e25331_app1.docx]

References

1. Copeland WE, Wolke D, Shanahan L, Costello EJ. Adult functional outcomes of common childhood psychiatric problems: A prospective, longitudinal study. JAMA Psychiatry 2015 Sep;72(9):892-899 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.0730] [Medline: 26176785]
2. Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. Dialogues Clin Neurosci 2009;11(1):7-20 [FREE Full text] [Medline: 19432384]
3. Gore FM, Bloem PJ, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: A systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]
4. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]
5. Contributing Lives, Thriving Communities – Report of the National Review of Mental Health Programmes and Services. Sydney, Australia: National Mental Health Commission, Australian Government; 2014. URL: https://www.mentalhealthcommission.gov.au/monitoring-and-reporting/national-reports/2014-contributing-lives-review [accessed 2021-06-14]
6. Rickwood DJ, Telford NR, Parker AG, Tanti CJ, McGorry PD. headspace - Australia's innovation in youth mental health: Who are the clients and why are they presenting? Med J Aust 2014 Feb 03;200(2):108-111. [doi: 10.5694/mja13.11235] [Medline: 24484115]
7. Hamilton BA, Naismith SL, Scott EM, Purcell S, Hickie IB. Disability is already pronounced in young people with early stages of affective disorders: Data from an early intervention service. J Affect Disord 2011 Jun;131(1-3):84-91. [doi: 10.1016/j.jad.2010.10.052] [Medline: 21112640]
8. O'Dea B, Glozier N, Purcell R, McGorry PD, Scott J, Feilds K, et al. A cross-sectional exploration of the clinical characteristics of disengaged (NEET) young people in primary mental healthcare. BMJ Open 2014 Dec 23;4(12):e006378 [FREE Full text] [doi: 10.1136/bmjopen-2014-006378] [Medline: 25537785]
9. Scott J, Scott EM, Hermens DF, Naismith SL, Guastella AJ, White D, et al. Functional impairment in adolescents and young adults with emerging mood disorders. Br J Psychiatry 2014 Nov;205(5):362-368. [doi: 10.1192/bjp.bp.113.134262] [Medline: 25213156]
10. Eiring O, Landmark BF, Aas E, Salkeld G, Nylenna M, Nytrøen K. What matters to patients? A systematic review of preferences for medication-associated outcomes in mental disorders. BMJ Open 2015 Apr 08;5(4):e007848 [FREE Full text] [doi: 10.1136/bmjopen-2015-007848] [Medline: 25854979]
11. Kilbourne AM, Fullerton C, Dausey D, Pincus HA, Hermann RC. A framework for measuring quality and promoting accountability across silos: The case of mental disorders and co-occurring conditions. Qual Saf Health Care 2010 Apr;19(2):113-116. [doi: 10.1136/qshc.2008.027706] [Medline: 20142404]
12. Hickie IB. Youth mental health: We know where we are and we can now say where we need to go next. Early Interv Psychiatry 2011 Feb;5 Suppl 1:63-69. [doi: 10.1111/j.1751-7893.2010.00243.x] [Medline: 21208394]

13. Iorfino F, Davenport TA, Ospina-Pinillos L, Hermens DF, Cross S, Burns J, et al. Using new and emerging technologies to identify and respond to suicidality among help-seeking young people: A cross-sectional study. J Med Internet Res 2017 Jul 12;19(7):e247 [FREE Full text] [doi: 10.2196/jmir.7897] [Medline: 28701290]
14. Scott EM, Hermens DF, Naismith SL, White D, Whitwell B, Guastella AJ, et al. Thoughts of death or suicidal ideation are common in young people aged 12 to 30 years presenting for mental health care. BMC Psychiatry 2012 Dec 26;12:234 [FREE Full text] [doi: 10.1186/1471-244X-12-234] [Medline: 23268688]
15. Iorfino F, Scott EM, Carpenter JS, Cross SP, Hermens DF, Killedar M, et al. Clinical stage transitions in persons aged 12 to 25 years presenting to early intervention mental health services with anxiety, mood, and psychotic disorders. JAMA Psychiatry 2019 Nov 01;76(11):1167-1175 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.2360] [Medline: 31461129]
16. Hickie IB, Scott EM, Cross SP, Iorfino F, Davenport TA, Guastella AJ, et al. Right care, first time: A highly personalised and measurement-based care model to manage youth mental health. Med J Aust 2019 Nov;211 Suppl 9:S3-S46. [doi: 10.5694/mja2.50383] [Medline: 31679171]
17. Woltmann E, Grogan-Kaylor A, Perron B, Georges H, Kilbourne AM, Bauer MS. Comparative effectiveness of collaborative chronic care models for mental health conditions across primary, specialty, and behavioral health care settings: Systematic review and meta-analysis. Am J Psychiatry 2012 Aug;169(8):790-804. [doi: 10.1176/appi.ajp.2012.11111616] [Medline: 22772364]
18. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002 Oct 09;288(14):1775-1779. [doi: 10.1001/jama.288.14.1775] [Medline: 12365965]
19. Raney L, Kathol R, Summergrad P. Collaborative care models for comorbid medical and behavioral health conditions. Focus (Am Psychiatr Publ) 2013 Jan;11(4):501-508. [doi: 10.1176/appi.focus.11.4.501]
20. Lora A, Lesage A, Pathare S, Levav I. Information for mental health systems: An instrument for policy-making and system service quality. Epidemiol Psychiatr Sci 2017 Aug;26(4):383-394 [FREE Full text] [doi: 10.1017/S2045796016000743] [Medline: 27780495]
21. Kilbourne AM, Beck K, Spaeth-Rublee B, Ramanuj P, O'Brien RW, Tomoyasu N, et al. Measuring and improving the quality of mental health care: A global perspective. World Psychiatry 2018 Feb;17(1):30-38 [FREE Full text] [doi: 10.1002/wps.20482] [Medline: 29352529]
22. Hickie IB, Davenport TA, Burns JM, Milton AC, Ospina-Pinillos L, Whittle L, et al. Project Synergy: Co-designing technology-enabled solutions for Australian mental health services reform. Med J Aust 2019 Oct;211 Suppl 7:S3-S39. [doi: 10.5694/mja2.50349] [Medline: 31587276]
23. Young HM, Nesbitt TS. Increasing the capacity of primary care through enabling technology. J Gen Intern Med 2017 Apr;32(4):398-403 [FREE Full text] [doi: 10.1007/s11606-016-3952-3] [Medline: 28243871]
24. Hoeft TJ, Fortney JC, Patel V, Unützer J. Task-sharing approaches to improve mental health care in rural and other low-resource settings: A systematic review. J Rural Health 2018 Dec;34(1):48-62 [FREE Full text] [doi: 10.1111/jrh.12229] [Medline: 28084667]
25. Falconer E, Kho D, Docherty JP. Use of technology for care coordination initiatives for patients with mental health issues: A systematic literature review. Neuropsychiatr Dis Treat 2018;14:2337-2349 [FREE Full text] [doi: 10.2147/NDT.S172810] [Medline: 30254446]
26. Kalton A, Falconer E, Docherty J, Alevras D, Brann D, Johnson K. Multi-agent-based simulation of a complex ecosystem of mental health care. J Med Syst 2016 Feb;40(2):39 [FREE Full text] [doi: 10.1007/s10916-015-0374-4] [Medline: 26590977]
27. Reger MA, Stanley IH, Joiner TE. Suicide mortality and coronavirus disease 2019-A perfect storm? JAMA Psychiatry 2020 Nov 01;77(11):1093-1094. [doi: 10.1001/jamapsychiatry.2020.1060] [Medline: 32275300]
28. Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/nejmp2008017]
29. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: Systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
30. Selby P, Vojtila L, Ashfaq I, Dragonetti R, Melamed OC, Carriere R, et al. Technology-enabled collaborative care for youth with early psychosis: A protocol for a feasibility study to improve physical health behaviours. Early Interv Psychiatry 2020 Aug 03:1. [doi: 10.1111/eip.13018] [Medline: 32748501]
31. Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016. Canberra, Australia: Australian Bureau of Statistics; 2018. URL: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012016?OpenDocument [accessed 2020-03-06]
32. General Population Health Needs Assessment. Coffs Harbour, Australia: North Coast Primary Health Network; 2018. URL: https://hnc.org.au/wp-content/uploads/2019/03/E.-General_Population_NeedsReport_Nov18.pdf [accessed 2021-06-14]
33. Census of Population and Housing: Estimating Homelessness 2016. Canberra, Australia: Australian Bureau of Statistics; 2018. URL: http://www.abs.gov.au/ausstats/abs@.nsf/mf/2049.0 [accessed 2020-03-06]
34. Occhipinti JA, Skinner A, Iorfino F, Lawson K, Sturgess J, Burgess W, et al. Reducing youth suicide: Systems modelling and simulation to guide targeted investments across the determinants. BMC Med 2021 Mar 12;19(1):61 [FREE Full text] [doi: 10.1186/s12916-021-01935-4] [Medline: 33706764]

35. Vemer P, Corro Ramos I, van Voorn GAK, Al MJ, Feenstra TL. AdViSHE: A validation-assessment tool of health-economic models for decision makers and model users. Pharmacoeconomics 2016 Apr;34(4):349-361 [FREE Full text] [doi: 10.1007/s40273-015-0327-2] [Medline: 26660529]
36. Badamgarav E, Weingarten SR, Henning JM, Knight K, Hasselblad V, Gano A, et al. Effectiveness of disease management programs in depression: A systematic review. Am J Psychiatry 2003 Dec;160(12):2080-2090. [doi: 10.1176/appi.ajp.160.12.2080] [Medline: 14638573]
37. Kikkert M, Goudriaan A, de Waal M, Peen J, Dekker J. Effectiveness of Integrated Dual Diagnosis Treatment (IDDT) in severe mental illness outpatients with a co-occurring substance use disorder. J Subst Abuse Treat 2018 Dec;95:35-42. [doi: 10.1016/j.jsat.2018.09.005] [Medline: 30352668]
38. van Os J, Guloksuz S, Vijn TW, Hafkenscheid A, Delespaul P. The evidence-based group-level symptom-reduction model as the organizing principle for mental health care: Time for change? World Psychiatry 2019 Feb;18(1):88-96 [FREE Full text] [doi: 10.1002/wps.20609] [Medline: 30600612]
39. Crowe J. Reform, revolution and disruption in mental health care: A consumer's perspective. Public Health Res Pract 2017 Apr 27;27(2):e2721711 [FREE Full text] [doi: 10.17061/phrp2721711] [Medline: 28474048]
40. Patrício L, Sangiorgi D, Mahr D, Čaić M, Kalantari S, Sundar S. Leveraging service design for healthcare transformation: Toward people-centered, integrated, and technology-enabled healthcare systems. J Serv Manage 2020 Jun 08;31(5):889-909. [doi: 10.1108/josm-11-2019-0332]
41. Hickie IB. The role of new technologies in monitoring the evolution of psychopathology and providing measurement-based care in young people. World Psychiatry 2020 Feb;19(1):38-39 [FREE Full text] [doi: 10.1002/wps.20697] [Medline: 31922686]
42. Subramanian S, Naimoli J, Matsubayashi T, Peters DH. Do we have the right models for scaling up health services to achieve the Millennium Development Goals? BMC Health Serv Res 2011 Dec 14;11:336 [FREE Full text] [doi: 10.1186/1472-6963-11-336] [Medline: 22168915]
43. Rowe SC, Davenport TA, Easton MA, Jackson TA, Melsness J, Ottavio A, et al. Co-designing the InnoWell Platform to deliver the right mental health care first time to regional youth. Aust J Rural Health 2020 Apr;28(2):190-194. [doi: 10.1111/ajr.12617] [Medline: 32281183]
44. Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. Psychother Res 2015;25(1):6-19. [doi: 10.1080/10503307.2013.817696] [Medline: 23885809]
45. Bradford S, Rickwood D. Electronic psychosocial assessment tool: Concept development and identification of barriers to successful implementation. J Technol Hum Serv 2014 Nov 26;32(4):275-296. [doi: 10.1080/15228835.2014.967906]
46. Kotte A, Hill KA, Mah AC, Korathu-Larson PA, Au JR, Izmirian S, et al. Facilitators and barriers of implementing a measurement feedback system in public youth mental health. Adm Policy Ment Health 2016 Nov;43(6):861-878. [doi: 10.1007/s10488-016-0729-2] [Medline: 27000148]

Abbreviations

AdViSHE: Assessment of the Validation Status of Health-Economic decision models
BMC: Brain and Mind Centre
ED: emergency department
K10: Kessler 10
NSW: New South Wales

Edited by C Basch; submitted 28.10.20; peer-reviewed by M Hamilton, A Hidki, X He; comments to author 01.12.20; revised version received 20.12.20; accepted 15.04.21; published 30.06.21.

Please cite as:
Iorfino F, Occhipinti JA, Skinner A, Davenport T, Rowe S, Prodan A, Sturgess J, Hickie IB
The Impact of Technology-Enabled Care Coordination in a Complex Mental Health System: A Local System Dynamics Model
J Med Internet Res 2021;23(6):e25331
URL: https://www.jmir.org/2021/6/e25331
doi:10.2196/25331
PMID:34077384

Original Paper

A Care Knowledge Management System Based on an Ontological Model of Caring for People With Dementia: Knowledge Representation and Development Study

Gyungha Kim[1,2*], PhD; Hwawoo Jeon[1,3*], MD; Sung Kee Park[1], PhD; Yong Suk Choi[3], PhD; Yoonseob Lim[1,4], PhD

[1]The Research Center for Diagnosis, Treatment and Care System of Dementia, Korea Institute of Science and Technology, Seoul, Republic of Korea
[2]Carbon & Light Materials Application Group, Korea Institute of Industrial Technology, Jeonju, Republic of Korea
[3]Department of Computer Science, Hanyang University, Seoul, Republic of Korea
[4]Department of HY-KIST Bio-convergence, Hanyang University, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Yoonseob Lim, PhD
The Research Center for Diagnosis, Treatment and Care System of Dementia
Korea Institute of Science and Technology
5, Hwarang-ro 14-gil, Seongbuk-gu
Seoul, KS013
Republic of Korea
Phone: 82 10 3700 3867
Email: yslim@kist.re.kr

Abstract

Background: Caregivers of people with dementia find it extremely difficult to choose the best care method because of complex environments and the variable symptoms of dementia. To alleviate this care burden, interventions have been proposed that use computer- or web-based applications. For example, an automatic diagnosis of the condition can improve the well-being of both the person with dementia and the caregiver. Other interventions support the individual with dementia in living independently.

Objective: The aim of this study was to develop an ontology-based care knowledge management system for people with dementia that will provide caregivers with a care guide suited to the environment and to the individual patient’s symptoms. This should also enable knowledge sharing among caregivers.

Methods: To build the care knowledge model, we reviewed existing ontologies that contain concepts and knowledge descriptions relating to the care of those with dementia, and we considered dementia care manuals. The basic concepts of the care ontology were confirmed by experts in Korea. To infer the different care methods required for the individual dementia patient, the reasoning rules as defined in Semantic Web Rule Languages and Prolog were utilized. The accuracy of the care knowledge in the ontological model and the usability of the proposed system were evaluated by using the Pellet reasoner and OntOlogy Pitfall Scanner!, and a survey and interviews were conducted with caregivers working in care centers in Korea.

Results: The care knowledge model contains six top-level concepts: care knowledge, task, assessment, person, environment, and medical knowledge. Based on this ontological model of dementia care, caregivers at a dementia care facility in Korea were able to access the care knowledge easily through a graphical user interface. The evaluation by the care experts showed that the system contained accurate care knowledge and a level of assessment comparable to normal assessment tools.

Conclusions: In this study, we developed a care knowledge system that can provide caregivers with care guides suited to individuals with dementia. We anticipate that the system could reduce the workload of caregivers.

(J Med Internet Res 2021;23(6):e25968) doi:10.2196/25968

KEYWORDS

caregiver; caregiver for person with dementia; knowledge model; ontology; knowledge management; semantic reasoning

Introduction

People with dementia exhibit memory loss, personality changes, depression, and instability in daily life, and they become highly dependent on family members and caregivers [1]. If a person has severe memory loss, family members must regularly check whether they are taking their medication, and they need to remind them of the detailed procedures of daily activities such as using the toilet or taking a shower. This high level of dependency demands constant attention and frequently requires additional support from care professionals. For this reason, most of the care costs relate to the need for family and social support rather than to the medical diagnosis [1]. Furthermore, the symptoms of dementia will differ depending on factors such as the living environment and personality [2]. The complexities of the care require a careful assessment of the physical and psychological abilities of the person with dementia and an assessment of the competence of those providing the care. One of the most difficult tasks for dementia caregivers is to choose the most effective approach for meeting the care needs and matching the behavioral characteristics of the individual [1,3-5].

To alleviate the care burden on caregivers, several studies have developed computer-based support systems that improve the well-being of the person with dementia [6-13]. These studies propose systems that support the caregivers by monitoring the health and daily activities of the person with dementia. The ambient-assisted intervention system (AAIS) offers ambient intelligence to improve quality of life by identifying the presence of behavioral and psychological symptoms of dementia (BPSDs) and by suggesting an appropriate intervention for the various symptoms [8]. For example, the AAIS will give repeated reminders to a person who frequently forgets to take their medicine. The K4Care project offers a computer-supported structure that represents the health care procedures needed to assist a person suffering from a particular disease, syndrome, or social issue [10]. In order to improve the quality of treatment that a person with dementia receives, decision support tools that help physicians detect wrong diagnoses, unobserved comorbidities, incomplete descriptions of the patient's condition, and appropriate prevention measures have been proposed [12,13]. For example, the Dementia Management and Support System (DMSS-R) supports the clinical routines and decision processes implemented by health professionals in their daily practice [12]. However, existing studies mainly focus on the clinical and statistical significance of the psychosocial and environmental interventions used to improve the welfare of persons with dementia [8-13].

Other studies have proposed interventions that can improve the independence of the person with dementia and so provide relief to caregivers [14-16]. A tablet-based app called MapHabit provides a customizable visual map that helps those with various Alzheimer-related disorders. For example, persons with memory loss can obtain step-by-step support in completing their daily activities by themselves while keeping connected with family members [14]. Another study introduced a digital memory notebook (DMN) application that is designed to help individuals with mild cognitive impairment to improve their everyday functioning [15]. The DMN is equipped with daily notes, schedule management, and to-do lists, and it helps patients to plan longer-term goals and to complete their everyday activities. Another study reviews the mobile apps that are designed to support the caregivers of older adults and that can be found on Google Play and iTunes [16]. These mobile apps generally address specific aspects such as information and resources, family communication, memory aids, care for the caregiver, behavioral solutions, medication management safety, or personal health record tracking [16].

On the whole, previous studies have attempted to enhance the physical and mental ability of the individual with dementia by proposing an intelligent system that either diagnoses their medical condition or increases their level of independence in daily activities. It is hoped that as the person with dementia becomes increasingly independent, less attention will be required from caregivers, and that this will eventually result in a reduced care burden for both family members and caregivers. However, since the BPSDs are highly complex and fluctuate sporadically, caregivers will still need to constantly monitor the health condition of the person with dementia and detect any changes in their behavioral symptoms [2,3]. This means that caregivers may have difficulty in choosing an appropriate care method regardless of the symptoms diagnosed previously. It has been suggested that to lessen the care burden on family members and caregivers, new approaches that are tailored to the patients, to the contexts where the symptoms occur, and to the caregivers are required [2,3,7].

In this study, we propose a care knowledge management system that can provide a care guide that is suited to the variable behavioral and psychological symptoms of the person with dementia by using an ontological knowledge model of dementia. The ontology-based model is a computer-interpretable knowledge model for formalizing and representing shared concepts in a specific domain of interest. This is expressed in a language that enables knowledge sharing among different applications and for various reasons [17-19]. The model can help improve data-driven decisions by explicitly defining and providing semantic concepts in a specific domain. Because of these advantages, the knowledge model for the proposed system could represent various care situations, which would include the symptoms of dementia. The Pellet and Prolog reasoner that will be built into the system can infer the proper care approach required for different individuals with dementia [20]. The graphical user interface (GUI) of the system will provide an easy way for family members or caregivers in care centers to share and update care data so that each caregiver can receive appropriate care guidelines simply by sending queries to the system without needing to ask the other caregivers or family members. To evaluate the accuracy of the care knowledge in the ontological model and the usability of the proposed system, we conducted a brief survey and interviews with caregivers working in care centers in Korea.

Methods

Determining the Domain and Scope of the Ontological Model of Dementia Care

To determine the domain and scope for the ontological knowledge model of dementia care, we tried to answer the following questions: (1) What is the target and scope to be covered by the ontology? (2) What will the ontology be used for? (3) How should the data be classified within the whole area of data available to users of the knowledge model? (4) How can the model represent knowledge to the target? (5) Who will be the users of the ontology? Based on the responses to these questions, we set the target for our knowledge model as a caregiver who actively cares for a person with mild dementia, who may be a family member, and we set the scope of the knowledge to include environmental information, schedule information, daily activity information, and disease information such as would be necessary to determine a care method for the individual with dementia.

Consideration of the Reuse of Existing Ontologies

We reviewed existing ontologies that contain various concepts and knowledge descriptions relevant to the care of persons with dementia. From the relevant ontologies, disease information relating to dementia was based on the AAIS [8] and K4Care approaches [10], and concepts involving environmental information were drawn from the KnowRob ontology [11].

Collecting Terms and Developing the Ontological Model of Dementia Care

We used two different methods to collect the terms and to define the classes for the ontological model: (1) a text-based document about caring for those with dementia and (2) interviews with care experts in the field of dementia. First, we collected the handbooks on dementia care that are already used in dementia care facilities in Korea [21-23]. From these care handbooks, we retrieved several sentences that describe the daily activities of people with dementia and related care methods. These care manuals also contain questions devised to evaluate the physical and psychological condition of the person with dementia and to assess their daily activities. From the sentences and questions collected, we extracted concepts relating to dementia care and then defined a top-level class with a hierarchy of concepts following a top-down approach [18]. A top-down development process starts with a definition of the most general concepts in the domain and is followed by subsequent specializations of the concepts. For example, we started with creating classes for the general concepts of “person,” “assessment,” “task,” “medical knowledge,” “environment,” and “care knowledge.” Then we specialized those classes further by creating their subclasses. The subclasses extracted were then arranged by mapping or by adding them to the class hierarchy. Overall, it took 3 months to build our initial version of the ontological model.

We then updated the terms and classes of the ontological model by interviewing 2 experts at a dementia center in Korea. Both experts had more than 10 years of experience in the dementia center, one as a director and the other as a manager. It took a total of 6 months to update the initial model, including 3 interviews with these experts. For example, we added relationships between the causes of the BPSDs in the “Medical Knowledge” class and the care methods in the “Care Knowledge” class and updated the decision rules based on the information provided during the interviews.

The Web Ontology Language (OWL) 2 language proposed by the World Wide Web Consortium was chosen to describe the ontology model [24]. The naming convention for class labels uses nouns and verbs with the first letter of a word being capitalized and multi-word labels (eg, Take_Meals and Take_Drug) having an underscore inserted between the words. Likewise, multi-word instance labels have an underscore inserted between the words and the instance (eg, Take_Meals_01 and Hypertension_Drug_01). The property names are set in the same way. The ontology was developed using the Protégé-OWL ontology editor [25]. Its consistency was checked using the Pellet reasoner [20] and the OntOlogy Pitfall Scanner (OOPS!) [26].

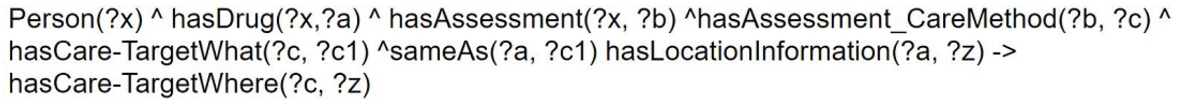
The two co-first authors were actively engaged in the entire process of building the ontological model. Both of them had previous experience in building an ontological model, and one of them has a PhD degree in ontology.

Decision Rules

We constructed reasoning rules using the OWL following Protégé, and the semantic rules were created using Semantic Web Rule Language (SWRL) and Prolog [24]. Reasoning rules in SWRL are easily defined within the Protégé-OWL ontology editor, but negative rules cannot be defined. We therefore decided to use SWRL for inferring the general care methods and Prolog for inferring individual care methods that reflect the behavior or psychological condition of a person with dementia. We followed actual care guidelines that we obtained either from interviews with care experts or from care manuals used in the dementia care center in Korea. Based on the information acquired, we designed the decision rules for caring for a person with dementia in different situations and when exhibiting different symptoms of dementia. For example, when retrieving the location of an object, the system utilizes reasoning rules that are defined in SWRL. However, for individual care methods, the system first infers general care methods using SWRL, and then the appropriate care methods for the individual are determined based on the reasoning rules in Prolog. Examples of SWRL rules are presented in Figures 1 and 2. The rules in Figure 1 are used to deduce the location of the hypertension drug for the patient. Person(?x) indicates that the variable x is a person. The binary relation hasDrug (?x, ?a) indicates that person x takes the drug a. If the person takes drug a and the location of the drug is known, the SWRL rule will provide the location of the drug for the person.

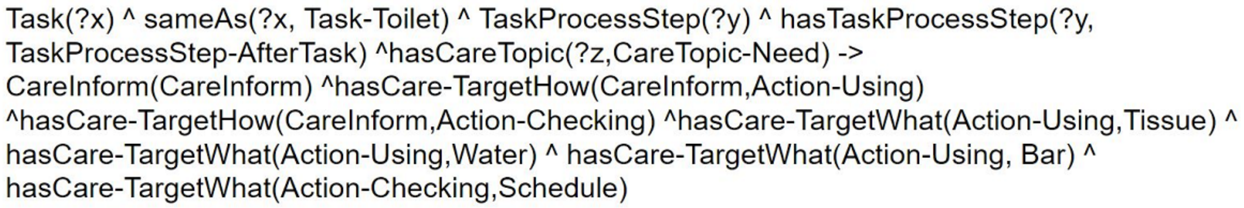
The rules in Figure 2 illustrate the process by which care guides direct the toilet use methods and procedures of the person with dementia. By following the rules, the system provides the inference result and informs the individual about the use of tissue, water, toilet rails, and schedule confirmation. The reasoning results from SWRL are refined by the Prolog reasoner, which utilizes the assessment information of the individual. For example, if the person with memory impairment frequently forgets to flush the toilet bowl, the Prolog reasoner will generate an additional care method that reminds the person with dementia to flush the water at the end of using the toilet.

Figure 1. Example of Semantic Web Rule Language rule about location notification.



```
Person(?x) ^ hasDrug(?x,?a) ^ hasAssessment(?x, ?b) ^hasAssessment_CareMethod(?b, ?c) ^
hasCare-TargetWhat(?c, ?c1) ^sameAs(?a, ?c1) hasLocationInformation(?a, ?z) ->
hasCare-TargetWhere(?c, ?z)
```

Figure 2. Example of Semantic Web Rule Language rule about Task-Toilet care.



```
Task(?x) ^ sameAs(?x, Task-Toilet) ^ TaskProcessStep(?y) ^ hasTaskProcessStep(?y,
TaskProcessStep-AfterTask) ^hasCareTopic(?z,CareTopic-Need) ->
CareInform(CareInform) ^hasCare-TargetHow(CareInform,Action-Using)
^hasCare-TargetHow(CareInform,Action-Checking) ^hasCare-TargetWhat(Action-Using,Tissue) ^
hasCare-TargetWhat(Action-Using,Water) ^ hasCare-TargetWhat(Action-Using, Bar) ^
hasCare-TargetWhat(Action-Checking,Schedule)
```

Evaluation of the Knowledge Management System

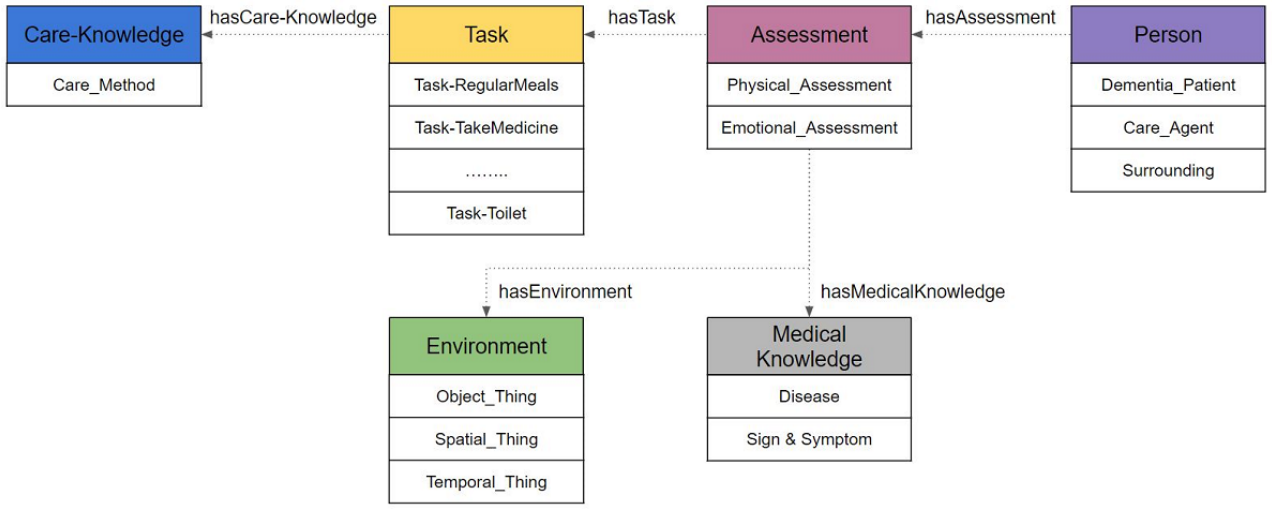
We first checked the care ontology with the OOPS! tool [26] to check its compliance with ontology authoring principles in addition to its structural quality. The knowledge model was also evaluated by care experts. Initially, we recruited 30 participants who each had more than a year of experience in caring for a person with dementia. However, only 4 of these participants had professional knowledge of dementia patient care, and we excluded the other 26 participants who had not been trained with professional nursing skills. The 4 participants selected had worked in the same dementia welfare center in Korea for an average of 10 years each, which included periods spent with other organizations. The average age of the participants was 60 years, with a range of 57 to 63 years. The respondents were all female (4/4, 100%). For evaluation, we provided the experts with a video introducing the knowledge management system we had built, followed by an opportunity to use the actual system. After watching the video, the experts were instructed to input patient information, and they then had an opportunity to check or to update the instances generated, such as the location of objects and the assessment data for the patient. A survey and an interview were conducted for 20 minutes after the actual experience with the knowledge management system. The whole evaluation took 30 minutes to an hour. Prior to the test, all participants gave their consent for the use of the experimental results, and the experiment was conducted with privacy protection (IRB number: K-2019-033). During the survey and interview, we asked the experts the following questions: (1) Is the ontology for the care of people with dementia consistent with the knowledge that the care expert holds? (2) Is the ontology for the care of people with dementia sufficient to determine their care guidelines? (3) Are the terms expressed in the ontology the same as the terms as used in actual care knowledge? (4) Is the knowledge management application convenient to use? (5) Does it seem useful for use in a real environment? The responses covered the accuracy, completeness, consistency, and usability of care knowledge on a 5-point scale (1=very invalid, 2=invalid, 3=moderate, 4=valid, and 5=very valid) [27-30]. The content validity index of the answers was calculated by taking the average.

Results

Care Ontology for People With Dementia

The ontological model for caring for people with dementia was based on six different top-level concepts achieved by classifying and defining information that is important in determining the care method needed for a person with dementia: person, task, environment, assessment, medical knowledge, and care knowledge. Figure 3 shows the overall structure of the top-level concepts and the relationships between them. The domain and range of each concept was defined, and the concepts were given properties. The proposed knowledge model has 502 classes and 230 properties.

Figure 3. Upper classes of the proposed care ontology for persons with dementia.



Person Class

The Person class represents the basic profile information on the person with dementia and their caregivers and family members. The basic profile information includes age, address, gender, and social relationship. The Person class was based on a previously developed ontology for supporting chronically ill patients and on the care handbook used in domestic dementia facilities in Korea [10].

Task Class

In order to provide a care method for the person with dementia, the activities of their daily life are defined as a Task class, and the Task class has a total of 20 different subtasks. These subtasks cover activities such as eating meals, sleeping, using the toilet, washing, taking medicine, and exercising. To describe the temporal and statistical aspects of each task, the Task class has the properties of task frequency, task execution cycle, task preference, task result, and task step.

Environment Class

The Environment class contains three different subclasses: Object_Thing, Spatial_Thing, and Temporal_Thing. Subclass Object_Thing consists of various objects in the living environment of the person with dementia. Subclass Spatial_Thing describes the location of an object, while temporal concepts such as time, birthday, and age are represented in the subclass Temporal_Thing. The structure of this class is based on the existing knowledge model KnowRob, which contains knowledge of various objects in the human environment [11].

Assessment Class

The Assessment class covers the behavioral competence of the person with dementia, including their physical ability to perform daily activities at home or in the care facilities and the psychological symptoms of their dementia. Most of the concepts in this class were adopted from the care manual used in welfare centers in Korea and from similar care handbooks [21-23]. The class is organized into the physical and psychological knowledge of the individual, and each instance of the class is associated with a care method described in the Care Knowledge class.

Medical Knowledge Class

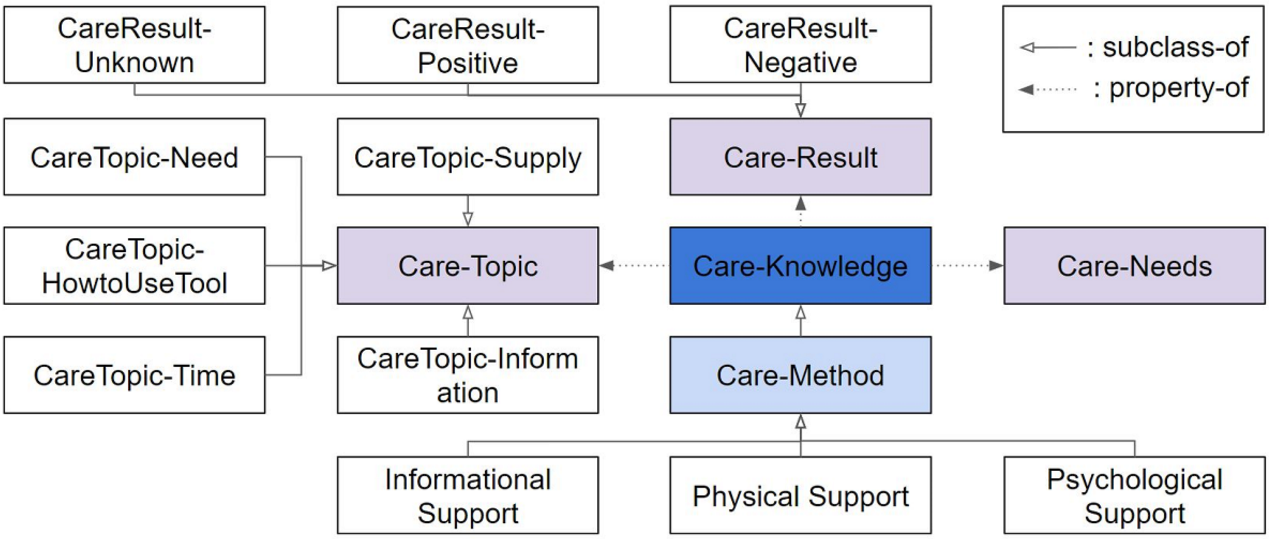
The Medical Knowledge class is a class that expresses the diseases and symptoms of dementia patients. The basic concepts and structure of this class were adopted from the K4Care ontology and the AAIS ontology [8,10].

Care Knowledge Class

The Care Knowledge class describes the various circumstances involved when helping the person with dementia, and most of the concepts defined in this class are linked with the Assessment, Environment, and Medical Knowledge classes.

The detailed structure of the Care Knowledge class is shown in Figure 4. The Care Knowledge class has three different subclasses: Informational Support, Physical Support, and Psychological Support. The Care Knowledge class has the properties of Care-Topic, Care-Needs, and Care-Result, which are used to describe the details of the care method and the history of the care results provided by the caregivers.

Figure 4. The structure of the “Care Knowledge” class in care ontology.



Care Knowledge Management System

We developed a care knowledge management system based on the knowledge model of caring for persons with dementia so that caregivers can access appropriate care guides and manage the care information through a GUI on personal computers. The overall architecture of the proposed system is shown in Figure 5. The knowledge management system can utilize cloud-based data such as weather and calendar information from Google. The context data manager handles the incoming queries from the user application and passes on the care knowledge inferred by the reasoner module. The reasoner is equipped with two different reasoning rules as described in the Methods section. The SWRL is defined for care ontology and is used for inferring general care methods. The Prolog reasoner infers individualized care methods for a particular patient based on the knowledge described in the assessment class that contains the physical and psychological conditions of the individual with dementia.

The care knowledge management system can be accessed via a GUI where the caregivers can manage or review instances of the ontology (Figure 6). For example, instances of Person class and Environment class can be accessed through the Profile and Location sections, respectively. To add new instances or to update the existing instances of assessment, the user needs to answer the list of questions corresponding to the selected topic (Care Method, Task, Body State, Psychological State, and Sociality). All the questions were created based on the actual assessment document used at the care center in Korea. Figures 6B and 6C show an example of questions on Care Method and instances of objects contained in the ontology.

Figure 5. Overall architecture of the care knowledge management system. OWL: Web Ontology Language; PwD: persons with dementia; SWRL: Semantic Web Rule Language.

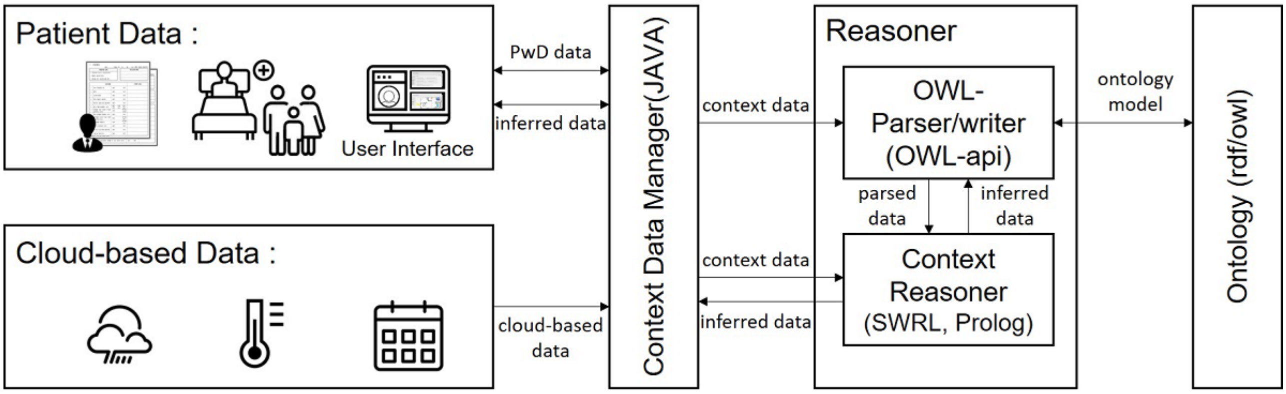
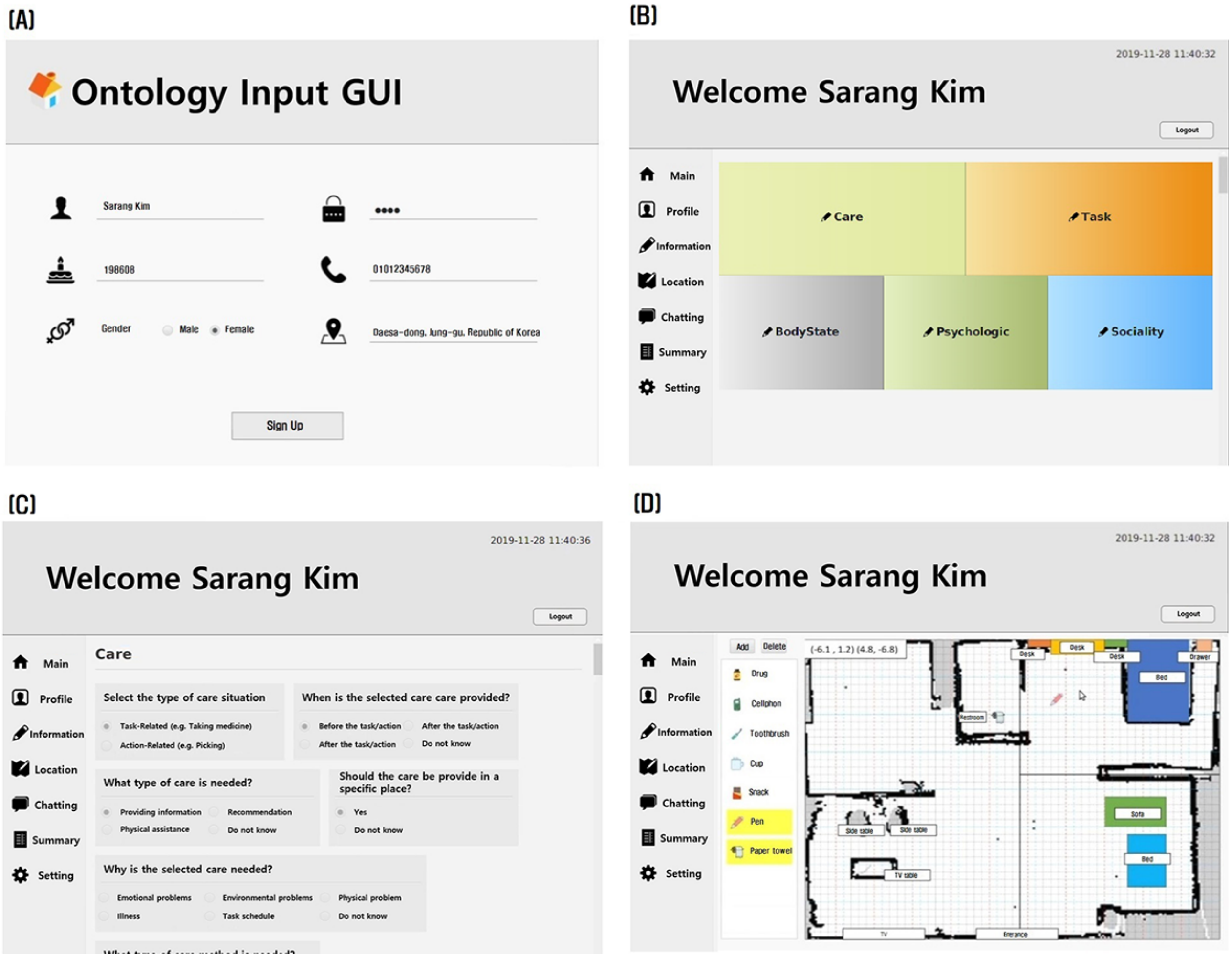


Figure 6. The graphical user interface design for care knowledge management: (A) an example of writing personal information on registration page, (B) selecting one of five different categories of assessment for persons with dementia, (C) updating assessment instances for care method, and (D) checking and updating the location of objects.



Sarang Kim

••••

198608

01012345678

Gender ○ Male ● Female

Daesa-dong, Jung-gu, Republic of Korea

Care

Select the type of care situation
● Task-Related (e.g. Taking medicine)
○ Action-Related (e.g. Picking)

When is the selected care care provided?
● Before the task/action ○ After the task/action
○ After the task/action ○ Do not know

What type of care is needed?
● Providing information ○ Recommendation
○ Physical assistance ○ Do not know

Should the care be provide in a specific place?
● Yes
○ Do not know

Why is the selected care needed?
○ Emotional problems ○ Environmental problems ○ Physical problem
○ Illness ○ Task schedule ○ Do not know

Case Study: Instances of Persons With Dementia

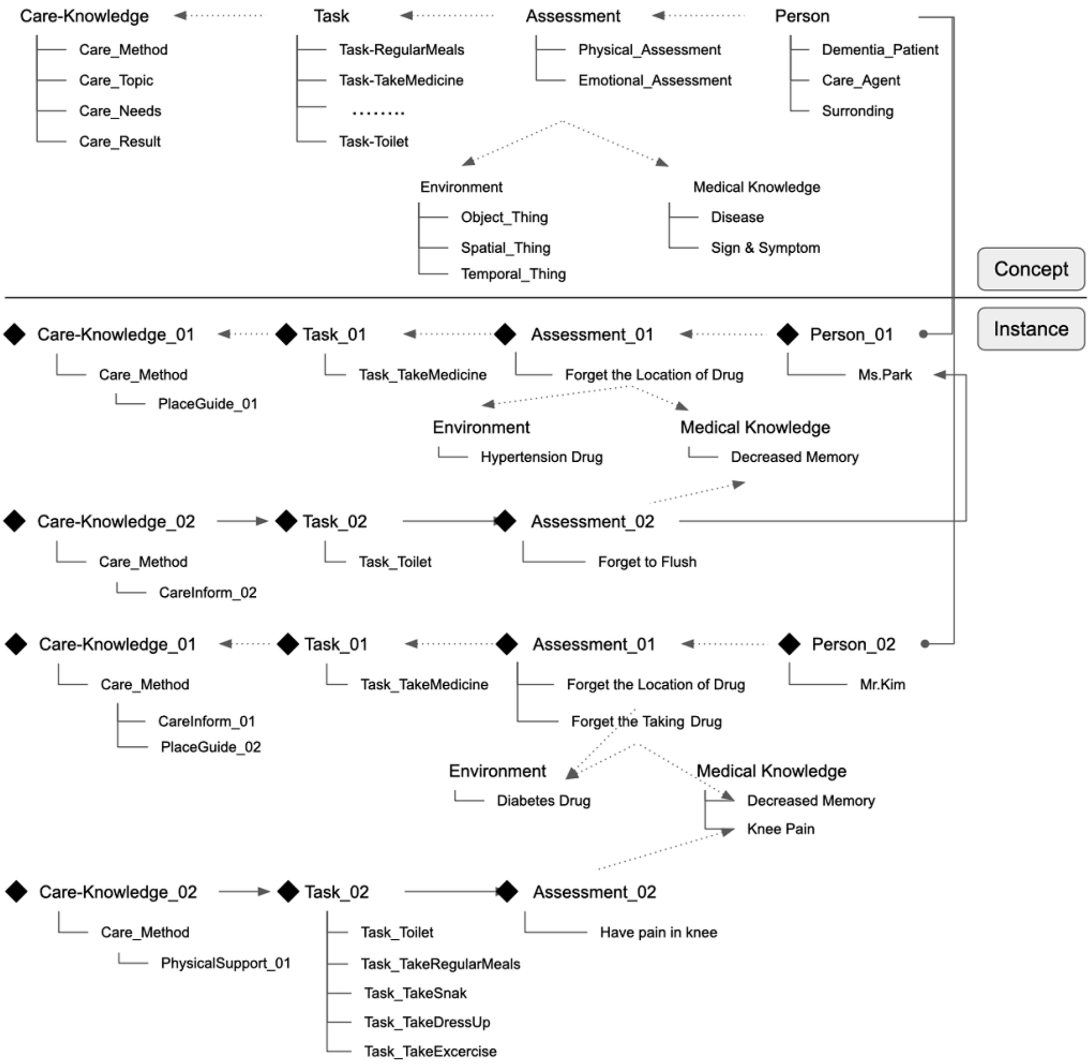
Table 1 presents a summary of the actual information for people with dementia written in a care manual obtained from a care center in Korea (we modified personal information such as name, address, and phone number). Ms Park has been diagnosed with mild dementia and is currently taking medication for chronic high blood pressure. She shows memory impairment symptoms such as not remembering the location of her drugs and the sequence of daily activities (eg, using the toilet). She can eat meals by herself but cannot control her appetite. The care manual also contains care guidelines for dealing with the individual's symptoms, such as helping her to remember an object's location and to flush the toilet bowl. Mr Kim has been diagnosed with mild dementia and is currently taking medication for diabetes. He has particular problems with memory impairment and regulation of his emotions.

Table 1. Assessment information on people with dementia obtained from the care center in Korea. Personal information such as name, gender, and address are modified.

Item	Example 1	Example 2
Name	Sooji Park	Inja Kim
Gender	Female	Male
Age (years)	70	80
Address	Jongno-gu, Seoul	Sungin-dong, Seoul
Contact	010-1234-0000	010-9876-0000
Dementia condition	Mild dementia	Mild dementia
Areas that need care	When she takes her high blood pressure medication, caregiver must remind her of the location of the medication. She eats well but cannot control her appetite. Caregiver should remind her to flush the toilet.	When he takes medicine, caregiver must remind him where the drug is and whether or not he has taken the medicine. He takes his diabetes drug three times a day after meals. When he moves to the bedroom, bathroom, or kitchen, he needs the caregiver's physical support. He cannot remember the number of meals taken.
Mental illness and symptoms	She has emotional anxiety. She has hallucinations and delusions.	He is emotionally unstable. He can suddenly become angry and sometimes throws objects violently.
Physical illness	She has high blood pressure.	He has diabetes and knee pain.
Activity	She goes to church every Sunday. She likes to clean her room.	He goes to church every weekend.
Others	She uses diapers at night. She uses dentures.	He uses dentures. A picture of his family helps to cheer him up.

We used the knowledge management system to add the information for Ms Park, and Figure 7 illustrates the instances of her data together with the structure of the upper concepts in the ontology. The individual's name, gender, age, address, and contact information are represented as instances of the Person class. Her mild dementia information, mental illness, and physical illness are represented in the Medical Knowledge class. Her task-related abilities, activities, and other precautions are expressed in the Assessment class. The care knowledge for informing on the location of the patient's hypertension drug is expressed as an instance, Assessment_01 with Care-Knowledge_01. The care knowledge of the patient's defecation activity is represented in Assessment_02 with Care-Knowledge_02. Care-Knowledge_02 thus expresses notification about using an object as a care method related to the defecation activity of the person with dementia.

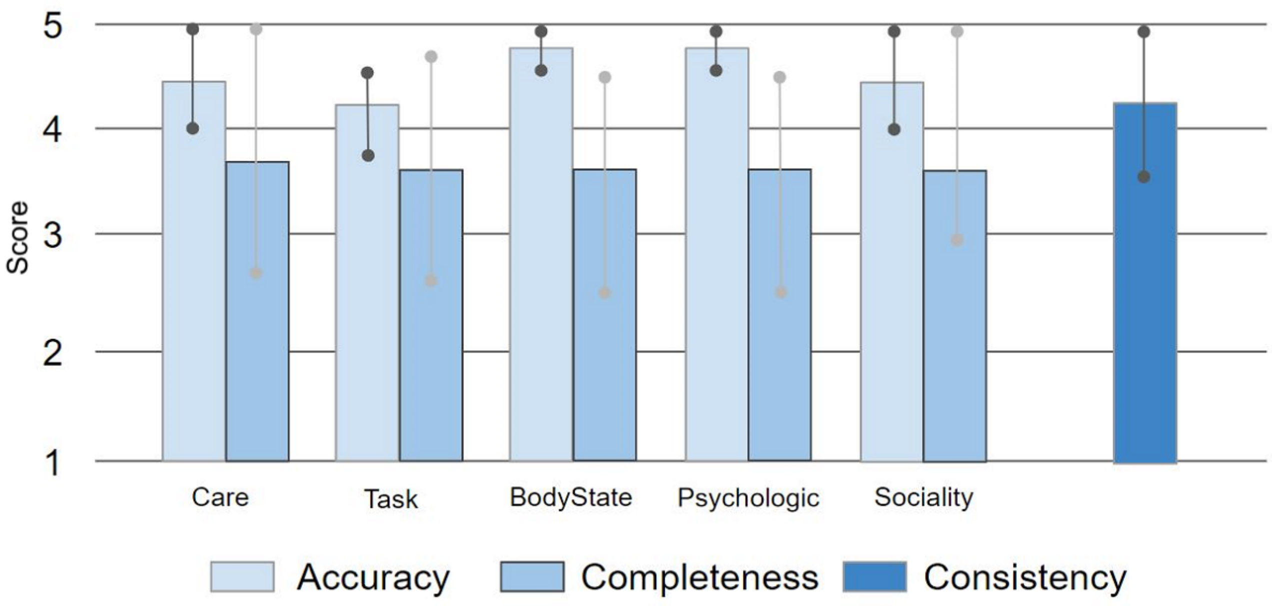
Figure 7. The structure of instances of a person with dementia together with upper-level concepts of care knowledge.



Evaluation of Care Knowledge Management System

We evaluated the care ontology with the OOPS! tool. We scored well on consistency, completeness, and conciseness, except for several minor issues such as a problem in determining the domain range and the property range for subclasses. In addition, we had our proposed system evaluated by 4 experts who had all been working in dementia care facilities for more than 10 years (the detailed procedure is described in the Methods section). The results of the evaluations received from experts are shown in Figure 8. To assess the accuracy of the ontology knowledge, we asked experts how accurately the proposed knowledge model described the actual knowledge used in a real care situation. In addition, in order to assess the completeness of the knowledge, we asked the experts whether the proposed system contained enough knowledge on caring for persons with dementia. In other words, the completeness score represents whether or not the knowledge model covers the necessary knowledge for caring for persons with dementia. Finally, regarding the consistency of the knowledge, the experts were asked whether or not the terms in the system developed were consistent with the terms used in the actual welfare centers. As a result, the caregivers evaluated that the current system contains accurate care knowledge (accuracy score: mean 4.55, SD 0.46) and covers most of the required care knowledge (completeness score: mean 3.80, SD 1.12). Regarding the consistency of the assessment information, the results show that the current ontology contains a comparable level of assessment information to the assessment tools (consistency score: mean 4.25, SD 0.83).

Figure 8. Evaluation of the care knowledge assessed by care experts in Korea.



Discussion

Principal Findings

In this study, we built an ontological knowledge model for dementia care and developed a care knowledge management system that can provide care guidelines tailored to various behavioral and psychological symptoms of dementia. The care knowledge management system includes specific activity types of the people with dementia with different dementia symptoms and care methods that are based on their dementia assessment. With the system, caregivers can easily get basic information on the different dementia patients that they care for and can be assisted with a care guide for the different tasks of the individuals.

It is well known that the symptoms of dementia are an important source of the burden and stress carried by caregivers. There is also evidence that exposure to chronic stress in the caregivers of those with dementia is associated with decreased physical health, mental illness, and poor quality of life in the caregiver [31-33]. Due to the high stresses of the care, caregivers often neglect or even abuse the person with dementia [4]. Numerous interventions have been proposed to change caregivers' behavior by increasing their social support and thereby improving their psychological state. There are home-based interventions, technology-driven interventions (ie, interventions that use computers or web-based applications), and interventions that are delivered via telephone. In a study by Godwin et al, 271 articles were reviewed for the efficacy of technology-driven interventions for the caregivers of persons with dementia [13]. They found that some of the studies showed reduced depression and anxiety among caregivers in the intervention group. However, the authors cautioned that no intervention method could completely reduce the burden on the caregiver because of the lack of technologies that monitored the patient's condition while supporting the caregiver.

Recent studies have attempted to provide better care service to persons with dementia by diagnosing their medical condition or by increasing their independence in daily activities. The DMSS-R supports the interventions performed by individual health professionals in daily practice with dementia patients [12]. DMSS-R is based on the Clinical Practice Guidelines for dementia diagnosis and provides diagnostic results, focusing on the physician's perspective on the patient. However, in addition to a clinical diagnosis of the dementia, caregivers need more information regarding the tasks of persons with dementia and their ability to perform these tasks. Other technologies, including a visual map and a memory notebook, help persons with dementia by presenting images relevant to their daily lives and schedules in context [14]. The visual map is a system that informs the individual of action guidelines through images and sound. It shows the necessary actions step by step through images that can help the person with memory impairment. However, there is a limit to the ability of persons with memory and behavioral impairments to live by themselves using only these systems without a caregiver's help. For example, an individual with dementia and physical disability could not be completely independent without the physical support of a caregiver, and this may inevitably require a care guide when making decisions.

We therefore focused on a situation where the caregivers could be relieved of remembering all the symptoms of the dementia patients they care for and of deciding on appropriate care methods for each one. To build a care knowledge system, we used the AAIS ontology and the K4Care ontology to obtain information on the disease types for dementia [8,10]. The AAIS ontology contains concepts of BPSDs that deal with the symptoms of impairment in perceptions, emotions, and behavior, while the K4Care ontology deals with the disease and the symptoms observed in the health care records of mild dementia patients. We included information about patients' physical diseases and social situations in addition to knowledge from the AAIS and K4Care ontologies, which included intervention information. We acquired object and spatiotemporal information from the KnowRob ontology to express knowledge regarding objects, times, and places relating to the person with dementia [11]. Additionally, the care ontology included individual care methods obtained from care-related books and the actual care guidelines used in dementia care facilities. Therefore, the proposed care ontology could represent not only the patient's disease symptoms, individual condition, and environmental information but also care methods suited to the different care situations.

To give caregivers easy access to the care knowledge and allow them to update information on the person with dementia, we developed a GUI for the knowledge management system (Figure 6). The GUI was designed for communication not only of personal information but also of symptoms, diseases, and social relationships that need to be considered when determining a care method. The system is able to manage and provide care-related knowledge for an unlimited number of persons with dementia and a large volume of additional care information. Although the main users of the proposed system would be professional caregivers in dementia care facilities, we believe that such a system could be easily extended so that nonprofessionals, such as family members, could access care information without special knowledge and that this would be beneficial by reducing both the burden and cost of caring for persons with dementia.

Limitations

Based on our interviews with professional caregivers, we realized that caregivers also pay considerable attention to emotional factors when deciding how to care for a person with dementia, and this explained the relatively low scores achieved by the current ontology for completeness (Figure 8). The proposed care ontology only considers the environmental factors, task performance, physical illness, mental illness, and social activities of the dementia patient. Due to the highly complex situations that can affect a person's emotional state, it is not easy to provide suitable care guidelines for the different emotional states of individuals with dementia. Future studies should investigate the range of psychological symptoms exhibited by persons with dementia with the relevant care that should be performed in care facilities so as to include the emotional dimension in future care knowledge. Various ontological models of human emotions could also be examined [34,35]. In addition, the developed care knowledge model targeted patients with mild dementia, and the experts who evaluated it cared for patients with mild dementia. Since the knowledge model deals only with knowledge related to mild dementia patients, it would be less helpful for the caregivers of patients with severe dementia.

The decision rules defined in the proposed knowledge management system are based on care methods described in interviews with care experts and in care manuals used in a care facility in Korea. The "Care Knowledge" and "Assessment" classes include terms and concepts that are described in these care manuals. For this reason, the care guidelines outlined in the proposed system would only be applicable to caregivers working in care facilities in Korea. Further studies are needed to extend the ontological model and the decision rules so that the proposed system would apply in places outside of Korea.

Finally, in this study, the accuracy and consistency of the care knowledge and the usability of the proposed knowledge system were evaluated by only 4 different care experts, although these experts each had an average of 10 years of dementia care experience. In addition, it remains to be tested whether the current system could reduce the care burden and help with care decision making in real care conditions. Currently, we are preparing a web-based application that can be accessed through mobile systems such as a tablet PC or smartphone for more convenient access to the care knowledge management system, and we plan to apply this updated system in real care situations.

Conclusions

In this study, we proposed a care knowledge management system that can provide the caregiver with detailed care information for the person with dementia. The system includes an ontological model of dementia care, a context data manager, and a rule-based reasoner in SWRL and Prolog. In addition, the system is equipped with a GUI, and it can be used anywhere in the home or welfare center regardless of the number or location of the dementia patients. Since the proposed system can provide care guidelines suitable for dementia patients, it is anticipated that the proposed system could reduce the workload of caregivers, but it remains to be tested in a future study.

Acknowledgments

We thank the caregivers working at the Seongbuk Welfare Center and the Gwangju Hanul Nursing Home for providing valuable feedback on the proposed system and helping us to improve it. This work was supported by the Technology Innovation Program (10077553, Development of Social Robot Intelligence for Social Human-Robot Interaction of Service Robots) and funded by the Ministry of Trade, Industry and Energy (Korea) and by a National Research Council of Science and Technology grant from the Korean government (Ministry of Science, ICT and Future Planning; grant CRC-15-04-KIST).

Conflicts of Interest

None declared.

References

1. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017 Dec 16;390(10113):2673-2734. [doi: 10.1016/S0140-6736(17)31363-6] [Medline: 28735855]

2. National Institute for Health and Care Excellence. Dementia: Supporting people with dementia and their carers in health and social care. URL: https://www.nice.org.uk/guidance/ng97 [accessed 2018-06-20]
3. Kales HC, Gitlin LN, Lyketsos CG. Assessment and management of behavioral and psychological symptoms of dementia. BMJ 2015 Mar 02;350:h369 [FREE Full text] [doi: 10.1136/bmj.h369] [Medline: 25731881]
4. Papastavrou E, Kalokerinou A, Papacostas SS, Tsangari H, Sourtzi P. Caring for a relative with dementia: family caregiver burden. J Adv Nurs 2007 Jun;58(5):446-457. [doi: 10.1111/j.1365-2648.2007.04250.x] [Medline: 17442030]
5. Villar-Taibo R, Martínez-Olmos MA, Bellido-Guerrero D, Calleja-Fernández A, Peinó-García R, Martís-Sueiro A, et al. Burden assessment in caregivers of patients with home artificial nutrition: a need and a challenge. Eur J Clin Nutr 2017 Feb;71(2):192-197. [doi: 10.1038/ejcn.2016.239] [Medline: 27901038]
6. 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]
7. Conger D, Rudnick C, Noland D, Newcomer M. Utilization of technology in Alzheimer's disease and mild cognitive impairment (MCI). Alzheimer's & Dementia 2019 Jul;15(7):P171-P172. [doi: 10.1016/j.jalz.2019.06.4351]
8. Navarro RF, Rodríguez MD, Favela J. Intervention tailoring in augmented cognition systems for elders with dementia. IEEE J Biomed Health Inform 2014 Jan;18(1):361-367. [doi: 10.1109/JBHI.2013.2267542] [Medline: 24403435]
9. Bastiani E, Librelotto GR, Freitas LO, Pereira R, Brasil MB. An Approach for Pervasive Homecare Environments Focused on Care of Patients with Dementia. Procedia Technology 2013;9:921-929. [doi: 10.1016/j.protcy.2013.12.103]
10. Riaño D, Real F, López-Vallverdú JA, Campana F, Ercolani S, Mecocci P, et al. An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. J Biomed Inform 2012 Jun;45(3):429-446 [FREE Full text] [doi: 10.1016/j.jbi.2011.12.008] [Medline: 22269224]
11. Tenorth M, Beetz M. KnowRob: A knowledge processing infrastructure for cognition-enabled robots. The International Journal of Robotics Research 2013 May 17;32(5):566-590. [doi: 10.1177/0278364913481635]
12. Lindgren H. Towards personalized decision support in the dementia domain based on clinical practice guidelines. User Model User-Adap Inter 2011 Jan 9;21(4-5):377-406 [FREE Full text] [doi: 10.1007/s11257-010-9090-4]
13. 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. [doi: 10.1177/1533317513481091] [Medline: 23528881]
14. Zola S, Golden M. P4-399: The use of visual maps as habit based assistive technology for individuals with Alzheimer's disease, Alzheimer's disease related dementias, and their caregivers. Alzheimer's & Dementia 2019 Jul 01;15:P1454-P1455. [doi: 10.1016/j.jalz.2019.06.4070]
15. Schmitter-Edgecombe M, Chudoba L, Church A, Dahmen J, Brown K. P3-009: A digital memory notebook application to support everyday functioning: Effectiveness and real-time intervention monitoring. Alzheimer's & Dementia 2019 Jul 01;15:P927-P928. [doi: 10.1016/j.jalz.2019.06.3035]
16. Grossman MR, Zak DK, Zelinski EM. Mobile Apps for Caregivers of Older Adults: Quantitative Content Analysis. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e162 [FREE Full text] [doi: 10.2196/mhealth.9345] [Medline: 30061093]
17. Noy NF. Semantic integration. SIGMOD Rec 2004 Dec 01;33(4):65. [doi: 10.1145/1041410.1041421]
18. Noy NF, McGuinness D. Ontology development 101: A guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880. URL: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf [accessed 2004-03-13]
19. Lasierra N, Roldán F, Alesanco A, García J. Towards improving usage and management of supplies in healthcare: An ontology-based solution for sharing knowledge. Expert Systems with Applications 2014 Oct;41(14):6261-6273. [doi: 10.1016/j.eswa.2014.04.023]
20. Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y. Pellet: A practical OWL-DL reasoner. Journal of Web Semantics 2007 Jun;5(2):51-53. [doi: 10.1016/j.websem.2007.03.004]
21. Cornish J. The dementia handbook: How to provide dementia care at home. Scotts Valley, CA: CreateSpace Independent Publishing Platform; 2017.
22. Monks KE. Home health nursing: Assessment and care planning, 4th edition. Maryland Heights, MO: Mosby; Oct 22, 2003.
23. Cantley C. A handbook of dementia care. Buckingham: Open University Press; 2001.
24. OWL 2 Web ontology language. URL: http://www.w3.org/TR/2009/REC-owl2-new-features-20091027/ [accessed 2021-05-04]
25. Noy NF, Sintek M, Decker S, Crubezy M, Fergerson R, Musen M. Creating Semantic Web contents with Protege-2000. IEEE Intell. Syst 2001 Mar;16(2):60-71. [doi: 10.1109/5254.920601]
26. Poveda-Villalón M, Gómez-Pérez A, Suárez-Figueroa MC. OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation. International Journal on Semantic Web and Information Systems 2014;10(2):7-34. [doi: 10.4018/ijswis.2014040102]
27. Gruninger M, Fox M. Methodology for the design and evaluation of ontologies. 1995 Apr 13 Presented at: the Workshop on Basic Ontological Issues in Knowledge Sharing; 1995; Montreal, Quebec.

28. Rao L, Osei-Bryson K. Towards defining dimensions of knowledge systems quality. Expert Systems with Applications 2007 Aug;33(2):368-378. [doi: 10.1016/j.eswa.2006.05.003]
29. Lee CY, Park HJ. A Case Study on Database Quality and Quality Factors. Journal of Information Technology Applications and Management 2004;11(4):209-225. [doi: 10.1016/j.eswa.2006.05.003]
30. Hong H. A Study on the Development of the Quality Evaluation Standard of Web-Based Databases. Journal of the Korean Society for Library and Information Science 2005 Jun 01;39(2):211-235. [doi: 10.4275/kslis.2005.39.2.211]
31. Rose-Rego SK, Strauss ME, Smyth KA. Differences in the perceived well-being of wives and husbands caring for persons with Alzheimer's disease. Gerontologist 1998 May;38(2):224-230. [doi: 10.1093/geront/38.2.224] [Medline: 9573667]
32. Clyburn LD, Stones MJ, Hadjistavropoulos T, Tuokko H. Predicting caregiver burden and depression in Alzheimer's disease. J Gerontol B Psychol Sci Soc Sci 2000 Jan;55(1):S2-13. [doi: 10.1093/geronb/55.1.s2] [Medline: 10728125]
33. Bell CM, Araki SS, Neumann PJ. The association between caregiver burden and caregiver health-related quality of life in Alzheimer disease. Alzheimer Dis Assoc Disord 2001;15(3):129-136. [doi: 10.1097/00002093-200107000-00004] [Medline: 11522930]
34. Gil R, Virgili-Gomá J, García R, Mason C. Emotions ontology for collaborative modelling and learning of emotional responses. Computers in Human Behavior 2015 Oct;51:610-617. [doi: 10.1016/j.chb.2014.11.100]
35. Abaalkhail R. Ontology based framework for conceptualizing human affective states and their influences.: School of Electrical Engineering and Computer Science, Faculty of Engineering, University of Ottawa; 2018. URL: https://ruor.uottawa.ca/handle/10393/38418 [accessed 2021-05-04]

Abbreviations

AAIS: ambient-assisted intervention system
BPSDs: behavioral and psychological symptoms of dementia
DMN: digital memory notebook
DMSS-R: Dementia Management and Support System
GUI: graphical user interface
OOPS!: OntOlogy Pitfall Scanner!
OWL: Web Ontology Language
SWRL: Semantic Web Rule Language

Edited by R Kukafka; submitted 25.11.20; peer-reviewed by R Rocha, L Kruesi; comments to author 20.01.21; revised version received 29.03.21; accepted 04.05.21; published 08.06.21.

Please cite as:
Kim G, Jeon H, Park SK, Choi YS, Lim Y
A Care Knowledge Management System Based on an Ontological Model of Caring for People With Dementia: Knowledge Representation and Development Study
J Med Internet Res 2021;23(6):e25968
URL: https://www.jmir.org/2021/6/e25968
doi:10.2196/25968
PMID:34100762

Original Paper

An Environmental Scan of Virtual "Walk-In" Clinics in Canada: Comparative Study

Spencer Matthewman[1], BSc; Sarah Spencer[1], MPH, MSc; M Ruth Lavergne[1], PhD; Rita K McCracken[2,3], MD, PhD; Lindsay Hedden[1,4], PhD

[1]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada
[2]Department of Family Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[3]Department of Family Medicine, Providence Health Care, Vancouver, BC, Canada
[4]British Columbia Academic Health Science Network, Vancouver, BC, Canada

Corresponding Author:
Lindsay Hedden, PhD
Faculty of Health Sciences
Simon Fraser University
8888 University Drive
Burnaby, BC, V5A 1S6
Canada
Phone: 1 236 521 207
Email: lindsay_hedden@sfu.ca

Abstract

Background: Canada has been slow to implement virtual care relative to other countries. However, in recent years, the availability of on-demand, "walk-in" virtual clinics has increased, with the COVID-19 pandemic contributing to the increased demand and provision of virtual care nationwide. Although virtual care facilitates access to physicians while maintaining physical distancing, there are concerns regarding the continuity and quality of care as well as equitable access. There is a paucity of research documenting the availability of virtual care in Canada, thus hampering the efforts to evaluate the impacts of its relatively rapid emergence on the broader health care system and on individual health.

Objective: We conducted a national environmental scan to determine the availability and scope of virtual walk-in clinics, cataloging the services they offer and whether they are operating through public or private payment.

Methods: We developed a power term and implemented a structured Google search to identify relevant clinics. From each clinic meeting our inclusion criteria, we abstracted data on the payment model, region of operation, services offered, and continuity of care. We compared clinics operating under different payment models using Fisher exact tests.

Results: We identified 18 virtual walk-in clinics. Of the 18 clinics, 10 (56%) provided some services under provincial public insurance, although 44% (8/18) operated on a fully private payment model while an additional 39% (7/18) charged patients out of pocket for some services. The most common supplemental services offered included dermatology (15/18, 83%), mental health services (14/18, 78%), and sexual health (11/18, 61%). Continuity, information sharing, or communication with the consumers' existing primary care providers were mentioned by 22% (4/18) of the clinics.

Conclusions: Virtual walk-in clinics have proliferated; however, concerns about equitable access, continuity of care, and diversion of physician workforce within these models highlight the importance of supporting virtual care options within the context of longitudinal primary care. More research is needed to support quality virtual care and understand its effects on patient and provider experiences and the overall health system utilization and costs.

(J Med Internet Res 2021;23(6):e27259) doi:10.2196/27259

KEYWORDS

virtual care; primary care; Canada; virtual health; patients; physicians

Introduction

Canada has lagged behind other countries in its uptake of virtual care as an integrated component of the health care system [1-3]. In 2018, less than 5% of physicians offered virtual services [2] despite most Canadians expressing their desire for video and phone visits as care options with the required technology being available.

The slow uptake prior to 2020 may be explained by multiple regulatory and physician compensation barriers, with few provinces providing a billing mechanism to physicians for virtual consultations [4]. The onset of the COVID-19 pandemic in Canada dramatically increased virtual care owing to the need for physical distancing to prevent disease transmission. This shift was supported by new or temporary billing codes and updated provincial mandates for providing virtual care where possible; in areas where provincial insurance plans, physician funding models, and limits on billable virtual visits had previously hindered physician uptake of virtual care, telephone or video consultations have now been adopted [5].

In addition to community-based physicians adapting their brick-and-mortar practices to include virtual care provision, the new fee codes for virtual visits provide an opportunity for the development of on-demand, “walk-in” virtual clinics that provide low-acuity, low-complexity care disconnected from existing physician-patient relationships. Some of these services are funded through provincial health insurance plans while others charge patients or are funded by supplemental private insurance plans directly.

Concerns have been raised that these services encourage episodic care, potentially contributing to fragmentation and poor continuity [1], and that they operate in a way that is not consistent with care that produces the best outcomes [6-10]. Additional research suggests that virtual walk-in services may have a detrimental effect on the quality of care [11,12], health care costs [13-15], and data privacy [11,12,15]. Furthermore, compared with virtual walk-in models, virtual care provided in the context of existing physician-patient relationships has proved more effective [7-9].

Despite these concerns, there is no existing research that catalogs the availability of these clinics or the services they offer. We conducted a national environmental scan to determine the availability and scope of virtual walk-in clinics offering synchronous appointments and prescriptions without the requirement or expectation of a longitudinal physician-patient relationship. We cataloged the services advertised by these clinics and determined whether they operated through public or private payment.

Methods

Approach

We used a structured Google search to identify virtual walk-in clinics across Canada, as we assumed this as one of the primary means that prospective patients would use to identify, locate, research, or connect with these services. We conducted a preliminary search on March 15, 2020, and a secondary search on June 2, 2020, to update and verify our initial results.

Keyword Search and Power Term

We compiled a list of words related to virtual health care and general medical care, from which we developed groups of search terms. Each term consisted of 3 words, with the first being Canada, the second relating to virtual care, and the final relating to the physician or clinic.

We analyzed 18 search terms for their strength in identifying virtual clinics. This process involved entering each term into Google, tallying the number of relevant sites listed, and analyzing 10 result pages per search term. Following this, we organized terms based on their strength; of the 18 search terms analyzed, the 6 strongest terms were selected. These were then combined using Boolean operators to form the following final search term: *Canada AND (virtual healthcare OR virtual health) AND (family medicine OR clinic OR general practitioner OR personalized care)*. We developed an initial list of clinics while developing our search strategy and then compared our final results with this initial list to ensure that the selected power term was displaying all the relevant sites.

Inclusion/Exclusion Criteria

To be included, the identified clinics had to meet the following criteria: (1) be based in Canada; (2) have a practicing medical doctor capable of remotely prescribing medication (ensuring all the services included for data extraction functioned as complete alternatives to traditional walk-in clinics and family physician appointments); (3) provide virtual visits through synchronous communication of some form (ie, phone, video, SMS text messaging); and (4) have English language websites. We excluded clinics that provided virtual services only to patients already enrolled with an associated brick-and-mortar clinic and those not providing primary care (eg, cancer clinics). Although such clinics provide care through virtual media, we felt their dependence on a preexisting physician-patient relationship and focus on specialist care largely differentiated them from their virtual walk-in counterparts.

Data Extraction

From each identified virtual walk-in clinic that met our inclusion criteria, we abstracted the following details from each site’s main pages and frequently asked questions sections and recorded them in a spreadsheet:

- virtual clinic name
- internet address
- geographic region(s) where services are available (select province[s] or all over Canada)
- enrollment type (membership, single visit, or both)
- source of payment (public provincial health insurance, private payment, or mixed)
- forms of synchronous communication offered (telephone, video, or SMS text messaging)
- use of artificial intelligence software to check symptoms and recommend treatments
- cost of membership and single visit

- examples of services offered (categorizing them iteratively, adding new categories when they were discovered, and then retroactively coding previous websites)

Additionally, we extracted free text that described data sharing or relationships with the patients' existing primary care physicians to investigate the extent to which these clinics are prioritizing continuity of care within the broader system and reflecting a prominent concern among health professionals regarding virtual walk-in clinics disrupting continuity [16]. Although we captured this information wherever available, we recognized that not all virtual clinics explicitly advertised their engagement with patients' existing physicians, and that in some cases, clinics may be limited in their ability to share patient information. Similarly, service listings are not expected to comprehensively represent everything offered; rather, they are a high-level indication of the type and scope of services available.

Analysis

We grouped similar services into service categories to streamline the data analysis. These categories included the following:

- specialist services (eg, oncology, endocrinology, pediatrics, obstetrics and gynecology, and sports medicine)
- individual behavior changes (eg, diet, weight loss, sleep therapy, smoking cessation, and problem gambling support)
- chronic disease management (eg, chronic obstetric pulmonary disease, diabetes, and heart failure)
- others (eg, hemorrhoid consultation, veterinarian consultation, emergency services, lactation consultation, and disability insurance claims/workers' compensation requests)

We grouped virtual clinics according to the mechanism of compensation for primary care services (publicly funded, privately funded, or mixed funding) as of June 2, 2020. Virtual clinics that bill the public health insurance plan in some provinces and patients directly in others were classified in the "mixed funding" category. We compared the availability, enrollment, communication type, and services offered across funding categories using Fisher exact tests (rather than chi-square tests owing to small cell counts). We have reported statistical significance using *P* values; however, as this analysis was descriptive, no decision rule with respect to statistical significance was applied. Analyses were conducted using Stata (StataCorp LLC) and Excel (Microsoft Corporation).

Ethics Statement

As all material gathered for data analysis was publicly available on the internet, no ethics approval was required.

Results

Search Results

We identified 19 virtual walk-in clinics during our initial March 15 search, 3 of which (Ontario Telemedicine Network [17], Novari Health [18], and Copeman Health [19]) were subsequently excluded owing to their requirement that clients already be associated with, or initiate care at, their brick-and-mortar clinics. We identified 2 additional services during our second search conducted on June 2, bringing the total to 18 walk-in clinic services that met our inclusion criteria.

Between the first and second searches, 2 virtual walk-in clinics added public reimbursement options. One of the services, Lumeca [20], switched from a national pay-for-use service to a completely publicly funded service available only to Saskatchewan residents. Similarly, Tia Health [21] added a public funding option for Alberta residents with a valid care card.

Source of Payment

More than half of the services we identified offered some form of public payment, with 5 being fully publicly funded and 5 operating on a public model in some provinces and a private one in others (Table 1). The other 8 clinics required patients to pay out of pocket for accessing services. Among pay-for-use services, the median cost per appointment was $41 ($32-$82). The median costs of memberships for individuals and families were $25 ($8-$29) and $41 ($12-$54) per month, respectively. The membership rates of 4 out of the 8 services were not made public on their websites.

Table 1. Services and fee structures categorized by compensation mechanisms.

Characteristic	Compensation mechanism, n (%)				*P* value
	Private (n=8, 44%)	Public (n=5, 28%)	Mix (n=5, 28%)	Total (N=18, 100%)	
Service availability					.002
National	7 (88)	0 (0)	5 (100)	12 (67)	
Provincial	1 (13)	3 (60)	0 (0)	4 (22)	
Multiprovincial	0 (0)	2 (40)	0 (0)	2 (11)	
Enrollment type					.21
Membership	5 (63)	4 (80)	1 (20)	10 (56)	
Single use	0 (0)	0 (0)	2 (40)	2 (11)	
Both	3 (38)	1 (20)	2 (40)	6 (33)	
Communication form					
Video call	7 (88)	5 (100)	5 (100)	17 (94)	.99
Telephone	4 (50)	3 (60)	3 (60)	10 (56)	.99
SMS text messaging	5 (63)	2 (40)	2 (40)	9 (50)	.71
Continuity with community general practitioners	2 (25)	2 (40)	0 (0)	4 (22)	.51

Virtual Walk-in Clinic Availability

Of the 18 walk-in clinics, 12 (67%) operated nationally, 4 (22%) within a single province, and 2 (11%) in multiple provinces. Privately funded services were more likely to be offered at the national level (7/8, 88%) than at the provincial (1/8, 13%) and multiprovincial (0/8, 0%) levels (*P*=.002). Of the 5 publicly funded services available, 3 were limited to single provinces, Saskatchewan and British Columbia, and 2 were offered to residents of multiple provinces. Although 5 nationally offered services had private and public payment options, they limited the public payment option to residents of British Columbia, Alberta, and Ontario having a valid health care card. Canadians using these services outside of these selected provinces would need to pay out of pocket.

Enrollment Type

Of the 18 services identified, 10 (56%) required membership, 2 (11%) were single-use services, and the remaining 6 (33%) offered both options. We observed that 63% (5/8) of the private payment clinics required membership, compared to 80% (4/5) of the public and 20% (1/5) of the mixed payment clinics (*P*=.21).

Communication Form

Possible communication forms consisted of video calls, telephone calls, and SMS text messaging. Video calls were the most common options across all clinics, offered by 88% (7/8) of the private payment clinics and 100% (5/5) of the mixed and publicly funded clinics. Among the 18 clinics, only 1 offered medical services and prescriptions without using video calls. This service, GOeVisit [18], instead relied on a medical form filled out by the patient and subsequently reviewed by a practitioner. Telephone and SMS text messaging options were still offered by most services but were less common than video calls.

Continuity of Care

Only 4 of the 18 services (22%) mentioned any form of continuity, information sharing, or communication with the consumers' existing primary care providers. Wello [22] and Akira [23], 2 privately funded services, mention an ability to work in tandem with their consumers' current practitioners and seeking access to the patients' current medical records; however, neither service explained how this was accomplished. Vivacare [24], a British Columbia–based clinic, offered continuity with physicians via their available in-person clinics (but not with physicians operating outside of their affiliated clinics). Lumeca [20], a publicly funded service for Saskatchewan residents, indicated that it would work with the family practitioners after obtaining written consent from the patients.

Telus' Babylon [25] and Tia Health [21] mentioned their ability to assign the same physician for each virtual appointment to help develop a patient-practitioner relationship; however, neither indicated the ability to support the existing relationships between patients and community-based primary care physicians, either by accessing the patients' existing records or by sharing the records of a virtual visit with the patients' regular primary care physicians. Of the 18 services, 3 explicitly stated that their services were not intended to replace a family doctor. Although other services may offer a form of information sharing without explicitly mentioning this on their websites, the results do suggest that continuity of care is not a primary concern for most virtual clinics.

Examples of Services Offered

Of the 18 virtual walk-in clinics included for data extraction, all except one—Outpost Health [26]—provided a list of their services (Table 2). The 3 most commonly mentioned services included skin care/dermatology, mental health services, and sexual health, mentioned by 83% (15/18), 78% (14/18), and 61% (11/18) of the clinics, respectively. Clinics using private payment were less likely to explicitly mention mental health services when compared with publicly funded and mixed payment clinics (*P*=.06).

Allergy treatment, support for individual behavior change, and specialist physician services were listed on the websites of more than half of the virtual care clinics. The remaining services offered were listed on the websites of one-third or fewer clinics. Clinics operating on a mixed funding model (private in some provinces and public in others) advertised the greatest breadth of health care services. It is also notable that none of the private payment clinics formally listed chronic disease management in their list of services, while 20% (1/5) of the public clinics and 60% of the mixed clinics (3/5) offered this service. One walk-in clinic—Teladoc [27]—listed emergency medical care among the services offered; this is in contrast with all other clinics, which explicitly stated that clients should visit a doctor or hospital immediately in the case of an emergency.

Table 2. Examples of offered services categorized by compensation mechanisms.

Services	Compensation mechanism, n (%)				*P* value
	Private (n=8, 44%)	Public (n=5, 28%)	Mix (n=5, 28%)	Total (N=18, 100%)	
Skin care/dermatology	5 (63)	5 (100)	5 (100)	15 (83)	.22
Mental health services	4 (50)	5 (100)	5 (100)	14 (78)	.06
Sexual and reproductive health	4 (50)	4 (80)	3 (60)	11 (61)	.82
Allergies	4 (50)	3 (60)	3 (60)	10 (56)	.99
Individual behavior change	3 (38)	2 (40)	5 (100)	10 (56)	.09
Specialist services	3 (38)	2 (40)	5 (100)	10 (56)	.09
Erectile dysfunction	1 (13)	1 (20)	4 (80)	6 (33)	.05
Chronic disease management	0 (0)	1 (20)	3 (60)	4 (22)	.04
Travel vaccinations	3 (38)	0 (0)	1 (20)	4 (22)	.51
Medical cannabis	1 (13)	1 (20)	1 (20)	3 (17)	.99
Naturopathy	1 (13)	0 (0)	2 (20)	3 (17)	.41
Other[a]	2 (25)	0 (0)	4 (80)	6 (33)	.03

[a] "Other" includes all services offered by <3 clinics, including the following: hemorrhoid consultation (2), veterinarian (2), emergency services (1), lactation consultation (1), and disability insurance claims/workers' compensation requests (1).

Discussion

Principal Results

Through a structured Google search, we identified 18 virtual walk-in clinics currently operating within Canada. This represents a 6-fold increase since 2015, when only 3 were available [28]. The rapid increase in the availability of these services coincides with the interest levels reported in existing consumer surveys [29] and the dramatic move toward virtual care delivery to facilitate physical distancing during the COVID-19 pandemic [5].

Virtual walk-in clinics provide a broad array of services, including primary care and other specialties, regardless of their payment models. Most clinics specifically advertised skin care, mental health, and sexual health services, supplementing basic primary care consultations, although less focus on mental health services was notable in privately funded clinics. Most relied on video calls as their primary means of communication; however, phone calls and SMS text messaging were also provided as communication options by most clinics.

Of the 18 virtual walk-in clinics that we identified, 15 charge patients out of pocket for core primary care services depending on the provinces where the patients resided. For example, 5 clinics were available nationally but limited public payment to patients with a valid health care card in British Columbia, Alberta, and Ontario. Conversely, 8 clinics operated on an entirely private payment model. Beyond the payment for primary care services, we noted that some clinics, namely Babylon [25] and Maple [30], also bundled supplemental services for which patients would pay out of pocket. Charging membership fees or out-of-pocket payments for physician consultations or services that are covered by provincial health insurance systems may contravene the *Canada Health Act* and provincial health insurance legislation [31], thus raising equity concerns. Additionally, although virtual walk-in clinics can offer an attractive work model for physicians—with predictable salaries and benefits, less overhead costs, and fewer administrative responsibilities [32,33]—this care model may attract physicians away from longitudinal community-based primary care services either partially or fully.

Few clinics reported that they facilitated communication or data sharing with patients' regular primary care providers (in either direction), suggesting that poor continuity of care may be a salient concern, thus reinforcing their suitability only for minor, less complicated conditions. The extensive use of virtual visits can potentially enhance access for patients who would normally face barriers when receiving primary care, such as people living in rural or remote areas, and those with compromised mobility and immune system challenges [34]; however, virtual walk-in clinics should not be used as substitutes for longitudinal relationships with primary care providers, particularly for patients with complex health and mobility issues.

Early evidence indicates that although the proportion of virtual visits has decreased since the lifting of pandemic restrictions, it has not returned to prepandemic levels. It remains to be seen where the balance between virtual care and in-person visits will settle following the pandemic; however, it is unlikely to return to prepandemic levels given the substantial federal investments and level of public demand [35]. Significant research is needed to address gaps in the knowledge on the quality of virtual episodic care and its effects on patient and provider experiences and overall health system utilization and costs.

Limitations

Our clinic searches were conducted in English only. This may have particularly resulted in the undercounting of the virtual walk-in clinics in Quebec. Second, our data extraction relied on the source material taken directly from clinic websites. Consequently, services not directly listed were not included. We may have underestimated the scope of services offered by some virtual walk-in services, as well as their potential information sharing with the patients' existing care providers. In future, clinics should be contacted directly to determine their service scope accurately and explore whether and how they receive information from or share information with community-based physicians. Third, our reliance on Google for executing the search strategy could have potentially underrepresented or missed clinics. Future research should involve multiple individuals and use additional search engines and virtual primary care service advertisements to strengthen the results. Lastly, given that only 18 nationwide virtual walk-in clinics were identified through our Google searches, our statistical analyses were hampered by the lack of statistical power.

Conclusions

This environmental scan sought to characterize the availability and scope of virtual walk-in clinics across Canada. We found a rapid increase in this care model, with 18 distinct services operating across the country, 15 of which required patients to pay out of pocket for some or all services offered. The implications of the rise in episodic virtual care could have negative effects on health care equity, quality, and costs; moreover, the growth of this model should be closely monitored and regulated by policy makers.

Availability of Data and Materials

The data that support the findings of this study are based on publicly available sources. The data set is available from the authors upon reasonable request.

Acknowledgments

This study was a student-led project and had no core funding. Funds for open access publication were provided by a Simon Fraser University Faculty Start-Up grant.

Conflicts of Interest

None declared.

References

1. Virtual care in Canada: Discussion paper. Canadian Medical Association. 2019 Aug. URL: https://www.cma.ca/sites/default/files/pdf/News/Virtual_Care_discussionpaper_v2EN.pdf [accessed 2020-09-21]
2. Connecting patients for better health: 2018. Canada Health Infoway. 2018. URL: https://tinyurl.com/27t5wscv [accessed 2020-09-21]
3. 2018 Canadian Physician Survey: Physicians' Use of Digital Health and Information Technologies in Practice. Canada Health Infoway. 2018 Dec. URL: https://infoway-inforoute.ca/en/component/edocman/3643-2018-canadian-physician-survey/view-document?Itemid=0 [accessed 2020-09-21]
4. Understanding the Use of and Compensation for Virtual-care Services in Primary Care. McMaster Health Forum. 2018 Jul 27. URL: https://tinyurl.com/57trpamw [accessed 2020-09-21]
5. Telemedicine and virtual care guidelines (and other clinical resources for COVID-19). Royal College of Physicians and Surgeons of Canada. 2020 Jul 06. URL: http://www.royalcollege.ca/rcsite/documents/about/covid-19-resources-telemedicine-virtual-care-e [accessed 2020-07-09]
6. Pereira Gray DJ, Sidaway-Lee K, White E, Thorne A, Evans PH. Continuity of care with doctors-a matter of life and death? A systematic review of continuity of care and mortality. BMJ Open 2018 Jun 28;8(6):e021161 [FREE Full text] [doi: 10.1136/bmjopen-2017-021161] [Medline: 29959146]
7. Shaw J, Jamieson T, Agarwal P, Griffin B, Wong I, Bhatia RS. Virtual care policy recommendations for patient-centred primary care: findings of a consensus policy dialogue using a nominal group technique. J Telemed Telecare 2018 Oct;24(9):608-615. [doi: 10.1177/1357633X17730444] [Medline: 28945161]
8. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ Open 2016 Jan 29;6(1):e009388 [FREE Full text] [doi: 10.1136/bmjopen-2015-009388] [Medline: 26826147]

9. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
10. Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, Hanson P, et al. Advantages and limitations of virtual online consultations in a NHS acute trust: the VOCAL mixed-methods study. Health Services and Delivery Research, No. 6.21. Southampton, UK: NIHR Journals Library; 2018 Jun. URL: https://www.ncbi.nlm.nih.gov/books/NBK507677/ [accessed 2020-09-22]
11. Quigley A, Hex N, Aznar C. Evaluation of Babylon: GP at hand. Ipsos MORI, York Health Economics Consortium. 2019 May. URL: https://www.hammersmithfulhamccg.nhs.uk/media/156123/Evaluation-of-Babylon-GP-at-Hand-Final-Report.pdf [accessed 2020-09-22]
12. Oliver D. David Oliver: Lessons from the Babylon Health saga. BMJ 2019 Jun 05;365:l2387. [doi: 10.1136/bmj.l2387] [Medline: 31167923]
13. McCracken R, Longhurst A, Lavergne R, Contadriopolous D. Virtual walk-in clinics undermine primary care. Policynote. 2019 Dec 19. URL: https://www.policynote.ca/virtual-primary-care/ [accessed 2020-10-03]
14. MacLeod A. How BC can fix primary health care, with or without corporations. The Tyee. 2020 Sep 11. URL: https://thetyee.ca/News/2020/09/11/How-BC-Can-Fix-Primary-Health-Care-Corporations/ [accessed 2020-10-04]
15. Hardcastle L, Ogbogu U. Virtual care: enhancing access or harming care? Healthc Manage Forum 2020 Nov;33(6):288-292 [FREE Full text] [doi: 10.1177/0840470420938818] [Medline: 32686506]
16. Glauser W. Virtual care has potential to fragment primary care and disturb continuity of care, warn doctors. CMAJ 2019 Sep 16;191(37):E1038-E1039 [FREE Full text] [doi: 10.1503/cmaj.1095796] [Medline: 31527195]
17. About OTN. Ontario Telemedicine Network. URL: https://www-origin.otn.ca/about/ [accessed 2020-11-24]
18. Novari eVisit - Virtual care/telemedicine. Novari Health. URL: https://www.novarihealth.com/evisit-telemedicine/ [accessed 2020-11-24]
19. 24/7 On-Demand Virtual Care Services By Copeman Healthcare Archived at webarchive.org. Copeman Healthcare. URL: https://web.archive.org/web/20200925170113/https://www.copemanhealthcare.com/private-healthcare/telehealth [accessed 2020-09-24]
20. Connect With a Doctor For Free, From Home. Lumeca. URL: https://lumeca.com/ [accessed 2020-11-25]
21. Tia Health. WELL Health Technologies Corporation. URL: https://tiahealth.com/ [accessed 2021-03-03]
22. FAQ/Support. Wello. URL: https://wello.ca/faq/ [accessed 2020-11-26]
23. Akira by TELUS Health. TELUS Health. URL: https://www.telus.com/en/health/organizations/group-health-benefits/employers/akira [accessed 2020-11-25]
24. Vivacare Telehealth. Tasly Vivacare Medical Group. URL: https://vivacare.ca/telehealth-online-doctor/ [accessed 2021-03-03]
25. Babylon by Telus Health. Telus Health. URL: https://web.archive.org/web/20210311081449/https://www.telus.com/en/bc/health/personal/babylon [accessed 2021-03-11]
26. Healthcare without Boundaries. Outpost Health. URL: https://www.outpost.health/ [accessed 2020-11-26]
27. Teladoc Health. Teladoc Health. URL: https://www.teladochealth.com/ [accessed 2021-03-04]
28. Liddy C, Hogel M, Blazkho V, Keely E. The current state of electronic consultation and electronic referral systems in Canada: an environmental scan. Stud Health Technol Inform 2015;209:75-83. [Medline: 25980708]
29. Connecting Patients for Better Health. Canada Health Infoway. 2018. URL: https://www.infoway-inforoute.ca/en/component/edocman/resources/reports/benefits-evaluation/3564-connecting-patients-for-better-health-2018 [accessed 2020-09-25]
30. Maple. URL: https://www.getmaple.ca [accessed 2021-03-04]
31. Government of Canada. Canada Health Act, RSC 1985, c. C-6, ss. 18-20. Justice Laws Website. URL: https://laws-lois.justice.gc.ca/eng/acts/c-6/page-1.html [accessed 2021-03-04]
32. Babylon Health - Babylon Family Physician (Vancouver Clinic) Archived at webarchive.org. Babylon Health. URL: https://web.archive.org/web/20210601220301/https://jobs.lever.co/babylonhealth/4a632742-2a04-4ae4-b131-1b6af70445b1 [accessed 2021-06-30]
33. Babylon health - join our team. Babylon Health. URL: https://www.babylonhealth.com/ca/physicians [accessed 2021-04-19]
34. Kelley LT, Phung M, Stamenova V, Fujioka J, Agarwal P, Onabajo N, et al. Exploring how virtual primary care visits affect patient burden of treatment. Int J Med Inform 2020 Sep;141:104228. [doi: 10.1016/j.ijmedinf.2020.104228] [Medline: 32683311]
35. Prime Minister announces virtual care and mental health tools for Canadians. Government of Canada. Ottawa, Ontario URL: https://pm.gc.ca/en/news/news-releases/2020/05/03/prime-minister-announces-virtual-care-and-mental-health-tools [accessed 2020-11-24]

Edited by R Kukafka; submitted 21.01.21; peer-reviewed by L Hardcastle, T Ong; comments to author 04.03.21; revised version received 05.05.21; accepted 06.05.21; published 11.06.21.

Please cite as:
Matthewman S, Spencer S, Lavergne MR, McCracken RK, Hedden L
An Environmental Scan of Virtual "Walk-In" Clinics in Canada: Comparative Study
J Med Internet Res 2021;23(6):e27259
URL: https://www.jmir.org/2021/6/e27259
doi:10.2196/27259
PMID:34114963

Letter to the Editor

The Numerous Benefits of Social Media for Medicine. Comment on “Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions”

John Alan Gambril[1,2], MD; Carter J Boyd[3], MBA, MD; Jamal Egbaria[4], MD

[1]Department of Internal Medicine, Ohio State University Wexner Medical Center, Columbus, OH, United States
[2]Nationwide Children's Hospital, Columbus, OH, United States
[3]Hansjörg Wyss Department of Plastic Surgery, NYU Langone Health, New York, NY, United States
[4]University of Alabama at Birmingham School of Medicine, Birmingham, AL, United States

Corresponding Author:
John Alan Gambril, MD
Department of Internal Medicine
Ohio State University Wexner Medical Center
410 W 10th Ave
Columbus, OH, 43210
United States
Phone: 1 205 936 2955
Email: alangambril@gmail.com

Related Article:

Comment on: https://www.jmir.org/2020/12/e25070/

(J Med Internet Res 2021;23(6):e27664) doi:10.2196/27664

KEYWORDS

social media; medical education; internet; academic medicine; promotion; tenure; health professions; scholarship; medicine; research; accomplishment; crowd source; contribution; innovation; education; dissemination

In their recent paper, Acquaviva et al [1] developed a set of guidelines to standardize curriculum vitae (CV) documentation of scholarly contributions made via social media platforms. Appropriately, the authors crowd-sourced contributions for the guidelines from the popular social media platform Twitter. Their work underscores the value of social media in idea sharing, highlights the growing role of online platforms in medical education, and signifies an important step in modernizing academic recognition to match the modernization of current medical learners.

Social media offers numerous scholarly and professional benefits [2,3]. These platforms have grown popular among the academic and medical communities as they are a means of networking with colleagues around the globe, discussing hot topics in various fields, engaging in medical education, sharing experiences through narrative medicine, and disseminating information to the lay-public [4]. More recently, social media has also taken on a vital role in residency recruitment. Without in-person interviews, programs have had to adopt new methods of sharing program strengths, highlighting unique program qualities, and appealing to applicants at an individual level [5].

We are avid proponents of academic social media. We can anecdotally attest to the educational value that arises from academic posts and discourse on social media platforms. Twitter, for example, offers a highly diverse pool of opinions covering all niches of medicine. It allows communication between individuals who might otherwise never interact. Sharing articles of interest via “tweeting” brings primary literature to your network’s fingertips. Twitter brings full professors and first-year medical students into the same arena of idea sharing. In a world that has increasingly recognized the shortcomings of traditional didactic lectures, social media offers modern educational methods better suited for today’s learners. Examples of this include “Tweetorials” (educational threads exploring a particular topic or phenomenon), podcasts, infographics, blogs, and virtual journal clubs among others [4].

These innovative methods of education through social media are intriguing to passionate educators. Those who seek to share knowledge and contribute to the advancement of scholarship will teach in whatever methods are most effective and will reach the most pupils. Importantly, social media is free, offering accessible medical education in a climate rife with expensive online materials and rising tuition. The value of these academic contributions must not go unrecognized. The time and dedication that goes into the development of educational posts through innovative methods should not be left out of an individuals' portfolio simply because the medium is not classic. If anything, the ingenuity and adaptability of the medium creates added value to the material. As the face of education evolves with our digital world, propelled forward by the COVID-19 pandemic, academia must evolve in unison to recognize these contributions.

We are thankful to Acquaviva and associates [1] for providing us with much-needed guidelines that allow for the documentation of education portfolios representative of today's evolving medical education environment.

Editorial Notice

The corresponding author of “Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions” declined to respond to this letter.

Conflicts of Interest

None declared.

References

1. Acquaviva KD, Mugele J, Abadilla N, Adamson T, Bernstein SL, Bhayani RK, et al. Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions. J Med Internet Res 2020 Dec 2;22(12):e25070 [FREE Full text] [doi: 10.2196/25070] [Medline: 33263554]
2. Spieler B, Ballard DH, Mazaheri P, Legro N, Catanzano T, Dey C, et al. Social Media in Radiology: Overview and Usefulness of Online Professional #SoMe Profiles. Acad Radiol 2021 Apr;28(4):526-539. [doi: 10.1016/j.acra.2020.03.014] [Medline: 32482471]
3. Mullins CH, Boyd CJ, Corey BL. Examining the Correlation Between Altmetric Score and Citations in the General Surgery Literature. J Surg Res 2020 Apr;248:159-164. [doi: 10.1016/j.jss.2019.11.008] [Medline: 31901796]
4. Sterling M, Leung P, Wright D, Bishop TF. The Use of Social Media in Graduate Medical Education: A Systematic Review. Acad Med 2017 Jul;92(7):1043-1056 [FREE Full text] [doi: 10.1097/ACM.0000000000001617] [Medline: 28225466]
5. Heitkamp NM, Morgan LE, Carmody JB, Heitkamp DE. Pediatric Program Directors Should Have an Active Presence on Twitter. Acad Pediatr 2021 Mar;21(2):201-204. [doi: 10.1016/j.acap.2020.11.016] [Medline: 33227535]

Abbreviations

CV: curriculum vitae

Edited by T Derrick; submitted 01.02.21; this is a non–peer-reviewed article;accepted 13.05.21; published 09.06.21.

Please cite as:
Gambril JA, Boyd CJ, Egbaria J
The Numerous Benefits of Social Media for Medicine. Comment on “Documenting Social Media Engagement as Scholarship: A New Model for Assessing Academic Accomplishment for the Health Professions”
J Med Internet Res 2021;23(6):e27664
URL: https://www.jmir.org/2021/6/e27664
doi:10.2196/27664
PMID:34106082

Letter to the Editor

Periodic Manual Algorithm Updates and Generalizability: A Developer's Response. Comment on "Evaluation of Four Artificial Intelligence–Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study"

Stephen Gilbert[1], BSc, BVMS, MRes, PhD; Matthew Fenech[1], MD, PhD; Anisa Idris[1], LLM, MPH, MBA; Ewelina Türk[1], MD

Ada Health, Berlin, Germany

Corresponding Author:
Stephen Gilbert, BSc, BVMS, MRes, PhD
Ada Health
Karl-Liebknecht-Str 1
Berlin, 10178
Germany
Phone: 49 017680396015
Email: stephen.gilbert@ada.com

Related Articles:

Comment on: https://www.jmir.org/2020/12/e18097

Comment in: https://www.jmir.org/2021/6/e29336

(J Med Internet Res 2021;23(6):e26514) doi:10.2196/26514

KEYWORDS

artificial intelligence; machine learning; mobile apps; medical diagnosis; mHealth; symptom assessment

We have several comments on the recent publication of Ćirković [1], in which repeated testing of four symptom assessment applications with clinical vignettes was carried out to look for "hints of 'non-locked learning algorithms'." As the developer of one of the symptom assessment applications studied by Ćirković [1], we are supportive of studies evaluating app performance; however, there are important limitations in the methodology of this study.

Most importantly, the methodology used in this study is not capable of addressing its main objective. The approach used to look for evidence of nonlocked algorithms was the quantification of differences in performance using 3 ophthalmology vignettes, first in 2018, then in 2020. This methodology, although highly limited due to the use of only 3 vignettes in one medical specialism, could be used to detect changes in app performance over time. It, however, cannot be used to distinguish between nonlocked algorithms and the manual updating of apps' medical intelligence, through the normal process of the manual release of updated app versions. Medical device regulations and quality system requirements provide standard mechanisms through which apps can be further developed, validated, and released as updated versions. The manual of medical knowledge in this manner has been acknowledged by the manufacturers of all the apps studied by Ćirković [1]. In response to previous independent vignettes studies [2,3], spokespeople for Your.MD and Babylon stated that they update their medical knowledge periodically, and this is also clear on Buoy's website. In Gilbert et al [4], the Ada app is described as having a knowledge base "built and reviewed by medical doctors in a curated process of knowledge integration from medical literature. It is being expanded continuously following this standardized process."

As is acknowledged in the limitations listed in Ćirković's work [1], the study used vignettes designed, entered, and with results adjudicated by a single clinician. This could result in bias and a narrow type of case. It is also acknowledged that 3 vignettes represent a small sample size for a vignettes study and that "standardized and transparent procedures" are needed for symptom assessment app–vignettes studies. We recently published a 200-vignette assessment of symptom assessment applications [4], including those studied by Ćirković [1], which used standardized and transparent procedures, including the separation of vignette design, entered and with results adjudication. It is our view that the effect of the limitations described by Ćirković [1], together with only including ophthalmological cases, is that the accuracy results reported have limited generalizability or repeatability. Our own internal validation testing shows an improvement in Ada's medical intelligence in all-condition top-3 suggestion accuracy (also known as M3, as defined by Miller et al [5]) of 4.8% between 2018 and 2020. We take account of all performance feedback we receive, and incorporate this, when judged appropriate by our medical knowledge experts, into updates of our app, through periodic releases of locked versions of our app.

Conflicts of Interest

All authors are employees of Ada Health.

References

1. Ćirković A. Evaluation of Four Artificial Intelligence-Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study. J Med Internet Res 2020 Dec 04;22(12):e18097 [FREE Full text] [doi: 10.2196/18097] [Medline: 33275113]
2. Burgess M. Can you really trust the medical apps on your phone? Wired. 2017 Oct 1. URL: https://www.wired.co.uk/article/health-apps-test-ada-yourmd-babylon-accuracy [accessed 2020-12-15]
3. What happened when Pulse tested symptom checker apps. Pulse. 2019 Aug 30. URL: http://www.pulsetoday.co.uk/news/analysis/what-happened-when-pulse-tested-symptom-checker-apps/20039333.article [accessed 2020-12-15]
4. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
5. Miller S, Gilbert S, Virani V, Wicks P. Patients' Utilization and Perception of an Artificial Intelligence-Based Symptom Assessment and Advice Technology in a British Primary Care Waiting Room: Exploratory Pilot Study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]

Edited by T Derrick; submitted 15.12.20; peer-reviewed by A Ćirković; accepted 13.05.21; published 16.06.21.

Please cite as:
Gilbert S, Fenech M, Idris A, Türk E
Periodic Manual Algorithm Updates and Generalizability: A Developer's Response. Comment on "Evaluation of Four Artificial Intelligence–Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study"
J Med Internet Res 2021;23(6):e26514
URL: https://www.jmir.org/2021/6/e26514
doi:10.2196/26514
PMID:34132641

Letter to the Editor

Author's Reply to: Periodic Manual Algorithm Updates and Generalizability: A Developer's Response. Comment on “Evaluation of Four Artificial Intelligence–Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study”

Aleksandar Ćirković, MD, MSc

Corresponding Author:
Aleksandar Ćirković, MD, MSc
Schulgasse 21
Weiden, 92637
Germany
Phone: 49 1788603753
Email: aleksandar.cirkovic@mailbox.org

Related Articles:

Comment on: https://www.jmir.org/2020/12/e18097

Comment on: https://www.jmir.org/2021/6/e26514

(J Med Internet Res 2021;23(6):e29336) doi:10.2196/29336

KEYWORDS

artificial intelligence; machine learning; mobile apps; medical diagnosis; mHealth; symptom assessment

I would like to thank Gilbert et al [1] for their useful contributions and tips to help improve our work [2].

Regarding their first point of critique, it is true that the evaluation lacked the technique to determine with certainty whether the algorithms used were or were not “locked” according to the terminology used by the Food and Drug Administration. For that reason, I stated in the *Discussion* section that some details of the study results only *suggest* a nonlocked direction [2]. Unfortunately, the abstract lacks this softening term. It remains true, however, that for regulators and physicians to correctly assess the functions of such software, either the algorithms would have to be disclosed—which might not be in the interest of the competing companies—or advanced testing models would have to be developed, possibly leading to a “cat-and-mouse” game similar to other regulatory fields. In developed countries where legal liabilities have to be clearly distributed and delimited, neither responsible physicians nor authorities will want to rely on mere published statements of company spokespeople alone—hard data from unbiased sources will be needed in the future.

The subsequent publication of the Ada study of a 200-vignette assessment is highly appreciated, as herein the authors were able to evaluate a considerably larger amount of data [3]. As I had stated, with a small sample size as a limitation, it does remain a possibility that the apps’ poor results as demonstrated in our study are within data variance norms or due to our own bias. However, the results did capture an experience that any potential user or patient with eye problems could have possibly encountered in the same way. Regulations will hopefully minimize risks for all users. Only thorough investigations, including but not limited to manufacturers’ evaluations, will help us better understand the effects the apps will have on public health. Until then, our knowledge base will consist of various analyses with possibly conflicting results that we will have to make sense of.

Authors' Contributions

The author AC is currently not affiliated with any institution, but is an Independent Scholar.

Conflicts of Interest

None declared.

References

1. Gilbert S, Fenech M, Idris A, Türk E. Periodic Manual Algorithm Updates and Generalizability: A Developer's Response. Comment on "Evaluation of Four Artificial Intelligence–Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study". J Med Internet Res 2021 Jun;23(6):e26514 [FREE Full text] [doi: 10.2169/26514]
2. Ćirković A. Evaluation of Four Artificial Intelligence-Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study. J Med Internet Res 2020 Dec 04;22(12):e18097 [FREE Full text] [doi: 10.2196/18097] [Medline: 33275113]
3. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]

Edited by T Derrick; submitted 02.04.21; this is a non–peer-reviewed article;accepted 13.05.21; published 16.06.21.

Please cite as:
Ćirković A
Author's Reply to: Periodic Manual Algorithm Updates and Generalizability: A Developer's Response. Comment on "Evaluation of Four Artificial Intelligence–Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study"
J Med Internet Res 2021;23(6):e29336
URL: https://www.jmir.org/2021/6/e29336
doi:10.2196/29336
PMID:34132643

Letter to the Editor

Bibliometric Studies and the Discipline of Social Media Mental Health Research. Comment on “Machine Learning for Mental Health in Social Media: Bibliometric Study”

Philip Resnik[1], PhD; Munmun De Choudhury[2], PhD; Katherine Musacchio Schafer[3], MEd, MS; Glen Coppersmith[4], PhD

[1]Department of Linguistics and Institute for Advanced Computer Studies, University of Maryland, College Park, MD, United States
[2]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States
[3]Department of Psychology, Florida State University, Tallahassee, FL, United States
[4]Qntfy, Boston, MA, United States

Corresponding Author:
Philip Resnik, PhD
Department of Linguistics and Institute for Advanced Computer Studies
University of Maryland
1401 Marie Mount Hall
College Park, MD, 20814
United States
Phone: 1 301 405 7002
Email: resnik@umd.edu

Related Articles:

Comment on: https://www.jmir.org/2021/3/e24870/

Comment in: https://www.jmir.org/2021/6/e29549

(J Med Internet Res 2021;23(6):e28990) doi:10.2196/28990

KEYWORDS

bibliometric analysis; machine learning; mental health; social media; bibliometrics

Bibliometric studies like the recent article by Kim et al [1] in the *Journal of Medical Internet Research* play an essential part in understanding the evolution of emerging, fast-moving research on machine learning for mental health in social media. However, the intended value of this paper’s contribution is tempered by some important lessons it teaches us about the current state of research on this topic.

The first key lesson is that computationally oriented research on mental health remains highly fragmented. Notably, variants on the cover term “mental health” are included in the illustrative search query but, crucially, “clinical psychology” and “psychiatry” are not. The terminological difference here reflects a prevailing technological focus often separated from clinical research and even more distant from clinical practice. Kim et al [1] do discuss a trend toward clinically validated self-report questionnaires to gather clinically relevant information. However, the review’s overall approach, from the search terms to the keyword analysis, simultaneously reflects and reinforces a widespread technological disregard for basic considerations in clinical psychology and psychiatry, such as the distinction between the symptoms of the disorders versus the disorders themselves. As technologists, we are often happy just to get our hands on enough data to work with. However, real progress toward solving these important problems demands a more careful definition of the actual mental health constructs under investigation and greater attention to the question of validity [2,3], with research questions and experimental choices guided by knowledge of the subject domain.

Second, the inclusion terms reflect a widespread narrow focus on methods, such as “neural network” and “hybrid intelligent system,” rather than the problems for which those methods are contributing solutions, such as “screening,” “risk assessment,” or “monitoring.” Even the cover term “natural language processing” focuses narrowly on engineering versus “computational linguistics” as a parent scientific discipline. Further lacking in the methodology-centric perspective are searches based on theoretical frameworks (which guide research, treatment, and intervention) or DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, 5th Edition) diagnoses (eg, major depressive disorder or persistent depressive disorder versus “depression,” which is not a diagnosis). The review reflects and reinforces a general tendency to frame machine learning research in terms of technical “tasks” rather than connecting them more directly with real-world problems, a necessary step toward translating technological progress into the broader mental health ecosystem within which the technology will ultimately need to be situated [4,5].

Third, the bibliometric approach taken here reflects a traditional top-down view that fails to break down information silos in a rapidly evolving field. It is now standard to cast the net more broadly by searching for citations in resources like Google Scholar and/or looking at papers’ references (cf Franklin et al [6]), and then narrow using exclusion criteria. Such practices can illuminate the wider space of relevant search terms and sources—for example, the notable absence of suicidality here among mental conditions, at least in the illustrative search—and uncover unexpected connections. Even within the most rigorous meta-analysis frameworks (Moher et al [7]), studies can miss “gray literature” (eg, conference proceedings, preprints, collected data that have never been analyzed, presented on, or published). For example, the substantially similar prior study by Chancellor and De Choudhury [3] needed to adjust for the limitations of indexing services, which had large gaps for conferences known to be important in this research area (eg, Association for the Advancement of Artificial Intelligence [AAAI], Association for Computational Linguistics [ACL], Association for Computing Machinery [ACM], Neural Information Processing Systems [NIPS/NeurIPS], American Medical Informatics Association [AMIA])—they were careful in particular to include the Workshop on Computational Linguistics and Clinical Psychology (CLPsych), a key interdisciplinary publication venue for natural language processing, machine learning, and mental health since 2014.

Kim et al [1] are to be commended for undertaking a bibliometric study with the goal of advancing our understanding of machine learning for mental health in social media. However, we would encourage thinking about their article as a different kind of contribution, even if not the intended one: it is an opportunity to draw attention to an increasing need, as the field grows, to approach this research space not only as technologists, but also as partners with clinical researchers and clinicians.

Conflicts of Interest

GC is a stockholder and employee of Qntfy.

References

1. Kim J, Lee D, Park E. Machine Learning for Mental Health in Social Media: Bibliometric Study. J Med Internet Res 2021 Mar 08;23(3):e24870 [FREE Full text] [doi: 10.2196/24870] [Medline: 33683209]
2. Ernala SK, Birnbaum ML, Michael L, Candan KA, Rizvi AF, Sterling WA, et al. Methodological Gaps in Predicting Mental Health States from Social Media: Triangulating Diagnostic Signals. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4-9; Glasgow, Scotland, UK p. 1-16. [doi: 10.1145/3290605.3300364]
3. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020;3:43 [FREE Full text] [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
4. Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim HC, et al. Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. Biol Psychiatry Cogn Neurosci Neuroimaging 2021 Feb 08. [doi: 10.1016/j.bpsc.2021.02.001] [Medline: 33571718]
5. Resnik P, Foreman A, Kuchuk M, Musacchio Schafer K, Pinkham B. Naturally occurring language as a source of evidence in suicide prevention. Suicide Life Threat Behav 2021 Feb;51(1):88-96. [doi: 10.1111/sltb.12674] [Medline: 32914479]
6. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. Psychol Bull 2017 Feb;143(2):187-232. [doi: 10.1037/bul0000084] [Medline: 27841450]
7. 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]

Abbreviations

AAAI: Association for the Advancement of Artificial Intelligence
ACL: Association for Computational Linguistics
ACM: Association for Computing Machinery
AMIA: American Medical Informatics Association
CLPsych: Computational Linguistics and Clinical Psychology
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
NIPS/NeurIPS: Neural Information Processing Systems

Edited by T Derrick; submitted 21.03.21; this is a non–peer-reviewed article;accepted 13.05.21; published 17.06.21.

Please cite as:
Resnik P, De Choudhury M, Musacchio Schafer K, Coppersmith G
Bibliometric Studies and the Discipline of Social Media Mental Health Research. Comment on "Machine Learning for Mental Health in Social Media: Bibliometric Study"
J Med Internet Res 2021;23(6):e28990
URL: https://www.jmir.org/2021/6/e28990
doi:10.2196/28990
PMID:34137722

Letter to the Editor

Authors' Reply to: Bibliometric Studies and the Discipline of Social Media Mental Health Research. Comment on "Machine Learning for Mental Health in Social Media: Bibliometric Study"

Jina Kim[1], BSc, MSc; Daeun Lee[1], BA; Eunil Park[1], BSc, MSc, PhD

Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul, Republic of Korea

Corresponding Author:
Eunil Park, BSc, MSc, PhD
Department of Applied Artificial Intelligence
Sungkyunkwan University
312 International Hall
Sungkyunkwan-ro 25-2
Seoul, 03063
Republic of Korea
Phone: 82 2 740 1864
Fax: 82 2 740 1856
Email: eunilpark@skku.edu

Related Articles:

Comment in: https://www.jmir.org/2021/6/e28990

Comment on: https://www.jmir.org/2021/3/e24870/

(J Med Internet Res 2021;23(6):e29549) doi:10.2196/29549

KEYWORDS

bibliometric analysis; machine learning; mental health; social media; bibliometrics

We thank all comments and critical insights on our study, "Machine Learning for Mental Health in Social Media: Bibliometric Study" [1]. In this letter, we responded to the discussion points raised by Resnik and colleagues [2].

First, Resnik et al [2] pointed out the limited scope of the search query selected in our bibliometric study, which concealed the approaches taken in the clinical research field. We agree that the search query may seem to be lacking in covering clinical research; however, we did not intend to distort the recent trends in machine learning for mental health in social media. Moreover, considering that publication venues including the *Journal of Medical Internet Research, BMJ Open, International Journal of Environmental Research and Public Health, Frontiers in Psychiatry*, and *Frontiers in Psychology* were listed as productive publication sources in the analysis results, we believe that the publications employed in our analysis cover the invaluable research methodologies in the medical research area. The completed list of retrieved publications can be found in the appendix [1].

Moreover, Resnik et al [2] also stated that the machine learning–related terms in the search query lean toward specific technological methodologies, which may not include solutions including screening, risk assessment, and monitoring. We could not agree more with their point of view on the importance of practical works applied to "real" clinical settings. However, since such approaches directly related to diagnosis or treatment required careful consideration, we had difficulties fully diving into it. Thus, we want to stress that the main focus of our research was to help clinical researchers and field affiliates understand the technical needs of machine learning, which we believe is necessary for the initiation phase of this research field. Besides, we conducted a thorough trend review analysis of highly cited papers to provide a more in-depth understanding of related research fields [3].

We appreciate the suggestions on keywords for the search query to include "clinical psychology," "psychiatry," and "computational linguistics." This led us to consider interviewing domain experts to set more domain-specific keywords in future research, following the procedure of the previous bibliometric study [4].

Third, Resnik and colleagues [2] raised concerns about only using indexing services and excluding search engines such as Google Scholar. Considering the rapidly changing trends in this research area, we employed both Web of Science and Scopus for the analysis to cover not only journal publications but conference proceedings. Nevertheless, we acknowledge Resnik and colleagues' concern and also mentioned the need to employ other medical-oriented databases, which may provide more significant academic and practical information in the *Discussion* section. We would also like to take this opportunity to recap the importance of nonpublished work by sharing a preprint paper in this research field [5], which helped us to take a step toward future research on applying machine learning and social media in mental health counseling.

We hope our analysis may trigger all stakeholders to further consider how to employ machine learning approaches toward mental health in social media and be a support to bridge the gap between technologists and clinicians.

Acknowledgments

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2020-0-01816) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2020R1C1C1004324).

Conflicts of Interest

None declared.

References

1. Kim J, Lee D, Park E. Machine Learning for Mental Health in Social Media: Bibliometric Study. J Med Internet Res 2021 Mar 08;23(3):e24870 [FREE Full text] [doi: 10.2196/24870] [Medline: 33683209]
2. Resnik P, De Choudhury M, Schafer Musacchio K, Coppersmith G. Bibliometric Studies and the Discipline of Social Media Mental Health Research. Comment on "Machine Learning for Mental Health in Social Media: Bibliometric Study". J Med Internet Res 2021 Jun;23(6):e28990. [doi: 10.2169/28990]
3. Wohlin C. An analysis of the most cited articles in software engineering journals—1999. Information and Software Technology 2005 Dec;47(15):957-964. [doi: 10.1016/j.infsof.2005.09.002]
4. Kalantari A, Kamsin A, Kamaruddin H, Ale Ebrahim N, Gani A, Ebrahimi A, et al. A bibliometric approach to tracking big data research trends. J Big Data 2017 Sep 29;4(1):1. [doi: 10.1186/s40537-017-0088-1]
5. Harrigian K, Aguirre C, Dredze M. On the State of Social Media Data for Mental Health Research. arXiv. Preprint posted online Nov 10, 2020 2021 [FREE Full text]

Edited by T Derrick; submitted 15.04.21; this is a non–peer-reviewed article;accepted 13.05.21; published 17.06.21.

Please cite as:
Kim J, Lee D, Park E
Authors' Reply to: Bibliometric Studies and the Discipline of Social Media Mental Health Research. Comment on "Machine Learning for Mental Health in Social Media: Bibliometric Study"
J Med Internet Res 2021;23(6):e29549
URL: https://www.jmir.org/2021/6/e29549
doi:10.2196/29549
PMID:34137721

Corrigenda and Addenda

Correction: Building Health Services in a Rapidly Changing Landscape: Lessons in Adaptive Leadership and Pivots in a COVID-19 Remote Monitoring Program

Celia Violet Laur[1*], MSc, PhD; Payal Agarwal[1,2*], MD; Geetha Mukerji[1,3], MD, MSc; Elaine Goulbourne[4], MSc; Hayley Baranek[1], MPH; Laura Pus[4], MBA; R Sacha Bhatia[1,5], MD, MBA; Danielle Martin[2,4,6], MD, MPP; Onil Bhattacharyya[1,2], MD, PhD

[1]Women's College Hospital Institute for Health System Solutions and Virtual Care, Toronto, ON, Canada
[2]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[3]Department of Medicine, Division of Endocrinology & Metabolism, University of Toronto, Toronto, ON, Canada
[4]Women's College Hospital, Toronto, ON, Canada
[5]University Health Network, Toronto, ON, Canada
[6]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[*]these authors contributed equally

Corresponding Author:
Onil Bhattacharyya, MD, PhD
Women's College Hospital Institute for Health System Solutions and Virtual Care
76 Grenville Street
Toronto, ON, M5S 1B2
Canada
Phone: 1 (416) 323 6400
Email: onil.bhattacharyya@wchospital.ca

Related Article:

Correction of: http://www.jmir.org/2021/1/e25507/

(J Med Internet Res 2021;23(6):e31044) doi:10.2196/31044

In “Building Health Services in a Rapidly Changing Landscape: Lessons in Adaptive Leadership and Pivots in a COVID-19 Remote Monitoring Program” (J Med Internet Res 2021;23(1):e25507), one error was noted.

In the originally published article, author Danielle Martin was associated with the incorrect ORCID number. In the corrected article, the ORCID number for this author has been updated as follows:

0000-0003-2517-2602

The correction will appear in the online version of the paper on the JMIR Publications website on June 10, 2021, 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 07.06.21; this is a non–peer-reviewed article;accepted 07.06.21; published 10.06.21.

Please cite as:
Laur CV, Agarwal P, Mukerji G, Goulbourne E, Baranek H, Pus L, Bhatia RS, Martin D, Bhattacharyya O
Correction: Building Health Services in a Rapidly Changing Landscape: Lessons in Adaptive Leadership and Pivots in a COVID-19 Remote Monitoring Program
J Med Internet Res 2021;23(6):e31044
URL: https://www.jmir.org/2021/6/e31044
doi:10.2196/31044
PMID:34111020

Corrigenda and Addenda

Correction: Rating Hospital Performance in China: Review of Publicly Available Measures and Development of a Ranking System

Shengjie Dong[1*], MPH; Ross Millar[2*], PhD; Chenshu Shi[3], MSc; Minye Dong[1], MPH; Yuyin Xiao[1], MPH; Jie Shen[4], MD; Guohong Li[1,4], PhD

[1]School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2]Health Services Management Centre, University of Birmingham, Birmingham, United Kingdom
[3]Center for Health Technology Assessment, China Hospital Development Institute, Shanghai Jiao Tong University, Shanghai, China
[4]China Hospital Development Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[*]these authors contributed equally

Corresponding Author:
Guohong Li, PhD
China Hospital Development Institute
Shanghai Jiao Tong University School of Medicine
227 South Chong Qing Road
Shanghai, 200025
China
Phone: 86 21 63846590
Email: guohongli@sjtu.edu.cn

Related Article:

Correction of: https://www.jmir.org/2021/6/e17095

(J Med Internet Res 2021;23(6):e31370) doi:10.2196/31370

In “Rating Hospital Performance in China: Review of Publicly Available Measures and Development of a Ranking System” (J Med Internet Res 2021;23(6):e17095) the authors noted one error.

An author was incorrectly included in the authorship list on the published paper. This author has been removed from the corrected version of the manuscript. The correct author list is Shengjie Dong, Ross Millar, Chenshu Shi, Minye Dong, Yuyin Xiao, Jie Shen, Guohong Li.

The correction will appear in the online version of the paper on the JMIR Publications website on June 22, 2021, 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 18.06.21; this is a non–peer-reviewed article;accepted 18.06.21; published 22.06.21.

Please cite as:
Dong S, Millar R, Shi C, Dong M, Xiao Y, Shen J, Li G
Correction: Rating Hospital Performance in China: Review of Publicly Available Measures and Development of a Ranking System
J Med Internet Res 2021;23(6):e31370
URL: https://www.jmir.org/2021/6/e31370/
doi:10.2196/31370
PMID:

Corrigenda and Addenda

Correction: Screening Tools: Their Intended Audiences and Purposes. Comment on “Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study”

Elizabeth Millen[1], BSc; Andreas Gilsdorf[1], MD, PhD; Matthew Fenech[1], MD, PhD; Stephen Gilbert[1], PhD

Ada Health, Berlin, Germany

Corresponding Author:
Elizabeth Millen, BSc
Ada Health
Karl-Liebknecht-Str 1
Berlin, 10178
Germany
Phone: 49 01520713083
Email: elizabeth.millen@ada.com

Related Article:

Correction of: https://www.jmir.org/2021/5/e26148

(J Med Internet Res 2021;23(6):e31268) doi:10.2196/31268

In “Screening Tools: Their Intended Audiences and Purposes. Comment on ‘Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study’” (J Med Internet Res 2021;23(5):e26148), one error was noted.

In the originally published manuscript, an incorrect statement was included in the Conflicts of Interest section:

> *EM, MF, and SG are employees of Ada Health GmbH. AG has no conflicts to declare.*

The statement has been corrected as follows:

> *All authors are employees of Ada Health GmbH.*

The correction will appear in the online version of the paper on the JMIR Publications website on June 24, 2021, 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 15.06.21; this is a non–peer-reviewed article;accepted 15.06.21; published 18.06.21.

Please cite as:
Millen E, Gilsdorf A, Fenech M, Gilbert S
Correction: Screening Tools: Their Intended Audiences and Purposes. Comment on “Diagnostic Accuracy of Web-Based COVID-19 Symptom Checkers: Comparison Study”
J Med Internet Res 2021;23(6):e31268
URL: https://www.jmir.org/2021/6/e31268/
doi:10.2196/31268
PMID:

Corrigenda and Addenda

Correction: Mobile Insight in Risk, Resilience, and Online Referral (MIRROR): Psychometric Evaluation of an Online Self-Help Test

Merel Marjolein van Herpen[1,2], MA; Manon A Boeschoten[3*], PhD; Hans te Brake[1*], PhD; Niels van der Aa[3], PhD; Miranda Olff[2,4], PhD

[1]ARQ Centre of Expertise for the Impact of Disasters and Crises, Diemen, Netherlands
[2]Department of Psychiatry, Amsterdam Neuroscience & Public Health, Amsterdam University Medical Center, Amsterdam, Netherlands
[3]ARQ Centre '45, Oegstgeest/Diemen, Netherlands
[4]ARQ National Psychotrauma Centre, Diemen, Netherlands
*these authors contributed equally

Corresponding Author:
Merel Marjolein van Herpen, MA
ARQ Centre of Expertise for the Impact of Disasters and Crises
Nienoord 5
Diemen,
Netherlands
Phone: 31 610082023
Email: m.van.herpen@impact.arq.org

Related Article:

Correction of: https://www.jmir.org/2020/9/e19716/

(J Med Internet Res 2021;23(6):e29776) doi:10.2196/29776

In “Mobile Insight in Risk, Resilience, and Online Referral (MIRROR): Psychometric Evaluation of an Online Self-Help Test” (J Med Internet Res 2020;22(9):e19716) one correction was made. This does not affect the analysis nor does it affect the interpretation or presentation of the results in the study.

In the “Methods” section, under “Measures” and “Depression, Anxiety and Stress”, the following phrase appears:

> *A 5-point response scale measures the extent to which each state has been experienced over the past week ranging from 0 (not at all) to 4 (most certainly).*

This has been replaced by the following:

> *A 4-point response scale measures the extent to which each state has been experienced over the past week ranging from 0 (did not apply to me at all) to 3 (applied to me very much, or most of the time).*

The correction will appear in the online version of the paper on the JMIR Publications website on June 4, 2021, 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 20.04.21; this is a non–peer-reviewed article;accepted 22.04.21; published 04.06.21.

Please cite as:
van Herpen MM, Boeschoten MA, te Brake H, van der Aa N, Olff M
Correction: Mobile Insight in Risk, Resilience, and Online Referral (MIRROR): Psychometric Evaluation of an Online Self-Help Test
J Med Internet Res 2021;23(6):e29776
URL: https://www.jmir.org/2021/6/e29776
doi:10.2196/29776
PMID:34086592

Corrigenda and Addenda

Correction: An Environmental Scan of Sex and Gender in Electronic Health Records: Analysis of Public Information Sources

Francis Lau[1*], PhD; Marcy Antonio[1*], BSc, MPH; Kelly Davison[1*], MSc, MN; Roz Queen[1*], BA; Katie Bryski[2], BA

[1]School of Health Information Science, University of Victoria, Victoria, BC, Canada
[2]Canada Health Infoway, Toronto, ON, Canada
[*]these authors contributed equally

Corresponding Author:
Francis Lau, PhD
School of Health Information Science
University of Victoria
P.O. Box 1700 STN CSC
Victoria, BC, V8W2Y2
Canada
Phone: 1 250 472 5131
Email: fylau@uvic.ca

Related Article:

Correction of: https://www.jmir.org/2020/11/e20050/

(J Med Internet Res 2021;23(6):e30764) doi:10.2196/30764

In “An Environmental Scan of Sex and Gender in Electronic Health Records: Analysis of Public Information Sources” (J Med Internet Res 2020;22(11):e20050) the authors noted three errors.

1. Under the “Results” section, second paragraph of “Types of Information Sources” subsection, the statement “*(3) OpenEHR and BioPortal as 2 international collaborations that have developed the Gender Archetype and Gender, Sex, and Sexual Orientation (GSSO) ontology, respectively;*” was published incorrectly. This has now been corrected as follows:

> *(3) OpenEHR as an international collaboration that has developed the Gender Archetype, and the BioPortal where the Gender, Sex, and Sexual Orientation (GSSO) ontology developed by Kronk [111] is hosted;*

2. Reference 111 was originally published as follows:

Gender, Sex and Sexual Orientation Ontology – Gender and Sex. BioPortal. URL: https://bioportal.bioontology.org/ontologies/GSSO/?p=classes&conceptid=root [accessed 2020-10-27]

This has been corrected to:

Kronk CA, Dexheimer JW. Development of the Gender, Sex, and Sexual Orientation ontology: Evaluation and workflow. J Am Med Inform Assoc 2020 Jul 01;27(7):1110-1115. [doi: 10.1093/jamia/ocaa061] [Medline: 32548638]

3. In the original article, the phone number of the corresponding author was published with a duplicate country code as *'1 12504725131'*.

This has been corrected to *'1 2504725131'*.

The correction will appear in the online version of the paper on the JMIR Publications website on June 4, 2021, 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 27.05.21; this is a non–peer-reviewed article;accepted 31.05.21; published 04.06.21.

Please cite as:
Lau F, Antonio M, Davison K, Queen R, Bryski K
Correction: An Environmental Scan of Sex and Gender in Electronic Health Records: Analysis of Public Information Sources
J Med Internet Res 2021;23(6):e30764
URL: https://www.jmir.org/2021/6/e30764
doi:10.2196/30764
PMID:34086590

Corrigenda and Addenda

Correction: Moderated Online Social Therapy for Young People With Active Suicidal Ideation: Qualitative Study

Eleanor Bailey[1,2,3], BA; Jo Robinson[1,3], BSc, MSc, PhD; Mario Alvarez-Jimenez[1,3], BSc, MResMeth, DClinPsy, PhD; Maja Nedeljkovic[2], MPH, MPsych(Clin), PhD; Lee Valentine[1,3], BA, MSocWk; Sarah Bendall[1,3], PGDipClinPsych, PhD; Simon D'Alfonso[1,4], BASc, PhD; Tamsyn Gilbertson[1,3], MSc, MPsych(Clin); Ben McKechnie[1], BA, MPsych(Clin); Simon Rice[1,3], MPsych(Clin), PhD

[1]Orygen, Parkville, Australia
[2]Swinburne University of Technology, Hawthorn, Australia
[3]Centre for Youth Mental Health, University of Melbourne, Parkville, Australia
[4]School of Computing and Information Systems, University of Melbourne, Parkville, Australia

Corresponding Author:
Eleanor Bailey, BA
Orygen
Locked Bag 10
35 Poplar Road
Parkville, 3052
Australia
Phone: 61 0412483600
Email: eleanor.bailey@orygen.org.au

Related Article:

Correction of: https://www.jmir.org/2021/4/e24260

(J Med Internet Res 2021;23(6):e29645) doi:10.2196/29645

In “Moderated Online Social Therapy for Young People With Active Suicidal Ideation: Qualitative Study” (J Med Internet Res 2021;23(4):e24260), the authors noted one error.

Incorrect qualifications were provided for author Ben McKechnie. The original qualifications provided were: BA, MPsych(Clin), MClinFamTher. However, MClinFamTher was provided in error and has been removed.

The correction will appear in the online version of the paper on the JMIR website on June 10, 2021, 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 15.04.21; this is a non–peer-reviewed article;accepted 03.06.21; published 10.06.21.

Please cite as:
Bailey E, Robinson J, Alvarez-Jimenez M, Nedeljkovic M, Valentine L, Bendall S, D'Alfonso S, Gilbertson T, McKechnie B, Rice S
Correction: Moderated Online Social Therapy for Young People With Active Suicidal Ideation: Qualitative Study
J Med Internet Res 2021;23(6):e29645
URL: https://www.jmir.org/2021/6/e29645
doi:10.2196/29645
PMID:34111019

Corrigenda and Addenda

Correction: Measurement of Digital Literacy Among Older Adults: Systematic Review

Sarah Soyeon Oh[1], PhD; Kyoung-A Kim[2], PhD; Minsu Kim[3]; Jaeuk Oh[3]; Sang Hui Chu[1], PhD; JiYeon Choi[1], PhD

[1]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea
[2]Department of Nursing, Yeoju Institute of Technology, Yeoju, Gyeonggi-do, Republic of Korea
[3]College of Nursing, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
JiYeon Choi, PhD
Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University
50 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3301
Email: jychoi610@yuhs.ac

Related Article:

Correction of: https://www.jmir.org/2021/2/e26145

(J Med Internet Res 2021;23(6):e30828) doi:10.2196/30828

In “Measurement of Digital Literacy Among Older Adults: Systematic Review” (J Med Internet Res 2021;23(2):e26145), the authors noted one error.

In the originally published paper, the funding statement appeared as follows:

> *This research was supported by the Brain Korea 21 FOUR Project funded by the National Research Foundation of Korea, Yonsei University College of Nursing. This study received funding from the National Research Foundation of Korea (grant number 2020R1A6A1A0304198911 [SO, KK, and JC]), the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A3A2067165 [SC]), and Yonsei University College of Nursing Faculty Research Fund (6-2020-0188 [SC and JC]).*

In the corrected version of the paper, the funding statement has been revised as follows:

> *This research was supported by the Brain Korea 21 FOUR Project funded by the National Research Foundation of Korea, Yonsei University College of Nursing. This study received funding from the National Research Foundation of Korea (grant number 2020R1A6A1A0304198911 [SO, KK, and JC]), the Ministry of Education of the Republic of Korea, and Yonsei University College of Nursing Faculty Research Fund (6-2020-0188 [SHC and JC]).*

The correction will appear in the online version of the paper on the JMIR Publications website on June 15, 2021, 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 31.05.21; this is a non–peer-reviewed article;accepted 31.05.21; published 15.06.21.

Please cite as:
Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J
Correction: Measurement of Digital Literacy Among Older Adults: Systematic Review
J Med Internet Res 2021;23(6):e30828
URL: https://www.jmir.org/2021/6/e30828
doi:10.2196/30828
PMID:34129513

Corrigenda and Addenda

Metadata Correction: Patient Empowerment During the COVID-19 Pandemic by Ensuring Safe and Fast Communication of Test Results: Implementation and Performance of a Tracking System

Gunnar Völkel[1*], PhD; Axel Fürstberger[1*], PhD; Julian D Schwab[1*], PhD; Silke D Werle[1*], PhD; Nensi Ikonomi[1*], MSc; Thomas Gscheidmeier[2], MSc; Johann M Kraus[1], PhD; Alexander Groß[1], PhD; Martin Holderried[3], MD, PhD; Julien Balig[1], MSc; Franz Jobst[4], MSc; Peter Kuhn[5], PhD; Klaus A Kuhn[6], PhD; Oliver Kohlbacher[7], PhD; Udo X Kaisers[4], MD; Thomas Seufferlein[8*], MD; Hans A Kestler[1*], PhD

[1]Institute of Medical Systems Biology, Ulm University, Ulm, Germany
[2]Department of Clinical Chemistry, University Hospital Ulm, Ulm, Germany
[3]Department of Medical Development and Quality Management, University Hospital Tübingen, Tübingen, Germany
[4]University Hospital Ulm, Ulm, Germany
[5]Comprehensive Cancer Center, University Hospital Ulm, Ulm, Germany
[6]Institute of Medical Informatics, Statistics and Epidemiology, Technical University of Munich, Ulm, Germany
[7]Institute for Translational Bioinformatics, University Hospital Tübingen, Tübingen, Germany
[8]Department of Internal Medicine I, University Hospital Ulm, Ulm, Germany
[*]these authors contributed equally

Corresponding Author:
Hans A Kestler, PhD
Institute of Medical Systems Biology
Ulm University
Albert-Einstein-Allee 11
Ulm, 89081
Germany
Phone: 49 731 500 24500
Fax: 49 731 500 24501
Email: hans.kestler@uni-ulm.de

Related Article:

Correction of: https://www.jmir.org/2021/6/e27348

(J Med Internet Res 2021;23(6):e31253) doi:10.2196/31253

In “Patient Empowerment During the COVID-19 Pandemic by Ensuring Safe and Fast Communication of Test Results: Implementation and Performance of a Tracking System” (J Med Internet Res 2021;23(6):e27348) the authors noted one error.

In the originally published manuscript, the author Nensi Ikonomi was not credited as an equal contributor. This has been corrected to show that Gunnar Völkel, Axel Fürstberger, Julian D Schwab, Silke D Werle, Nensi Ikonomi, Thomas Seufferlein, and Hans A Kestler all contributed equally to the manuscript.

The correction will appear in the online version of the paper on the JMIR Publications website on June 21, 2021, 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 15.06.21; this is a non–peer-reviewed article;accepted 16.06.21; published 21.06.21.

Please cite as:
Völkel G, Fürstberger A, Schwab JD, Werle SD, Ikonomi N, Gscheidmeier T, Kraus JM, Groß A, Holderried M, Balig J, Jobst F, Kuhn P, Kuhn KA, Kohlbacher O, Kaisers UX, Seufferlein T, Kestler HA
Metadata Correction: Patient Empowerment During the COVID-19 Pandemic by Ensuring Safe and Fast Communication of Test Results: Implementation and Performance of a Tracking System
J Med Internet Res 2021;23(6):e31253
URL: https://www.jmir.org/2021/6/e31253
doi:10.2196/31253
PMID:34152993

Original Paper

COVID-19 Knowledge Resource Categorization and Tracking: Conceptual Framework Study

Muhammad Afzal[1], PhD; Maqbool Hussain[1], PhD; Jamil Hussain[2], PhD; Jaehun Bang[3], PhD; Sungyoung Lee[3], PhD

[1]Department of Software, Sejong University, 209 Neungdong-ro, Seoul, Republic of Korea
[2]Department of Data Science, Sejong University, Seoul, Republic of Korea
[3]Department of Computer Science and Engineering, Kyung Hee University, Yongin, Republic of Korea

Corresponding Author:
Muhammad Afzal, PhD
Department of Software
Sejong University
209 Neungdong-ro
Innovation Center
Office 422
Seoul, 05006
Republic of Korea
Phone: 82 234082973
Email: mafzal@sejong.ac.kr

Abstract

Background: Since the declaration of COVID-19 as a global pandemic by the World Health Organization, the disease has gained momentum with every passing day. Various private and government sectors of different countries allocated funding for research in multiple capacities. A significant portion of efforts has been devoted to information technology and service infrastructure development, including research on developing intelligent models and techniques for alerts, monitoring, early diagnosis, prevention, and other relevant services. As a result, many information resources have been created globally and are available for use. However, a defined structure to organize these resources into categories based on the nature and origin of the data is lacking.

Objective: This study aims to organize COVID-19 information resources into a well-defined structure to facilitate the easy identification of a resource, tracking information workflows, and to provide a guide for a contextual dashboard design and development.

Methods: A sequence of action research was performed that involved a review of COVID-19 efforts and initiatives on a global scale during the year 2020. Data were collected according to the defined structure of primary, secondary, and tertiary categories. Various techniques for descriptive statistical analysis were employed to gain insights into the data to help develop a conceptual framework to organize resources and track interactions between different resources.

Results: Investigating diverse information at the primary, secondary, and tertiary levels enabled us to develop a conceptual framework for COVID-19–related efforts and initiatives. The framework of resource categorization provides a gateway to access global initiatives with enriched metadata, and assists users in tracking the workflow of tertiary, secondary, and primary resources with relationships between various fragments of information. The results demonstrated mapping initiatives at the tertiary level to secondary level and then to the primary level to reach firsthand data, research, and trials.

Conclusions: Adopting the proposed three-level structure allows for a consistent organization and management of existing COVID-19 knowledge resources and provides a roadmap for classifying future resources. This study is one of the earliest studies to introduce an infrastructure for locating and placing the right information at the right place. By implementing the proposed framework according to the stated guidelines, this study allows for the development of applications such as interactive dashboards to facilitate the contextual identification and tracking of interdependent COVID-19 knowledge resources.

(J Med Internet Res 2021;23(6):e29730) doi:10.2196/29730

KEYWORDS

information organization; resource management; knowledge graphs; interactive dashboard; dependency tracking; COVID-19; pandemic; information technology; tracing information; dashboards; digital health

Introduction

The novel coronavirus—SARS-CoV-2—first appeared in December 2019 and has quickly spread over other regions of the world. The World Health Organization (WHO) declared COVID-19, the disease caused by SARS-CoV-2, a global pandemic [1,2]. It gained momentum as every day passed, and private and government sectors of different countries pushed funding toward research on COVID-19 in various capacities. A portion was dedicated to investing in vaccine discovery and personal protective equipment manufacturing. The other portion was devoted to information technology and service infrastructure development, including research to develop intelligent models and techniques for alerts, diagnosis, treatment, prognosis, prevention, and other relevant services [3]. Our study focused on the second part of the research initiatives related to COVID-19 and provides a comprehensive review of such initiatives.

A rapid and timely response was required from the world to circumvent the challenge of COVID-19. Various organizations from different countries concentrated on investigations and discoveries in data, information, and knowledge to support population health [4]. We designed a structure for categorizing efforts and initiatives related to COVID-19 into three levels or categories: primary, secondary, and tertiary. The primary level represents resources and initiatives that accrue raw data or research about patients with COVID-19 and bring it into the global space. The secondary level encompasses resources and initiatives that analyze primary-level resources, making it more meaningful by adding metadata and filtering out unnecessary items in the data. The tertiary level includes resources and initiatives that consolidate the first- and second-level efforts by creating guidelines, code systems, standard resources, and vocabularies. This study provides a review of initiatives globally at these three levels and proposes a novel research dashboard for tracking COVID-19 resources with dependency workflows. The proposed dashboard will provide a gateway to global initiatives with enriched contextual metadata to help users track the information flow at different levels for validation and verification.

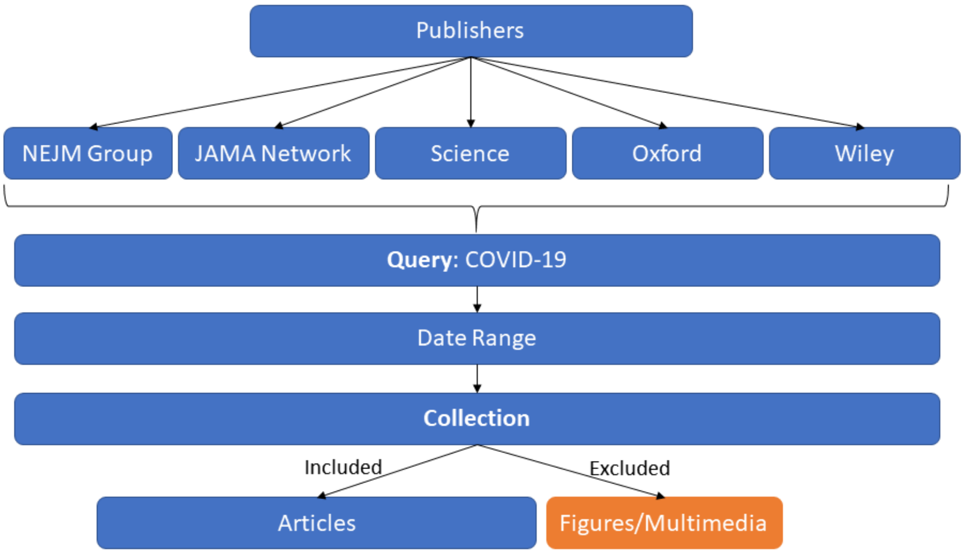
This study is the first of its kind to categorize efforts and initiatives in the form of resources related to COVID-19. It was inspired from the research sources outlined in commentaries and library handouts we consulted [5,6]. This study improvises the three types of information sources to map COVID-19–related initiatives to each kind and devises a method of finding their interdependence for easy tracking of information in contextual trails and metadata. In the context of COVID-19, we can find studies that discuss data sets and techniques, but no such research has been found to categorize initiatives and their dependencies.

Methods

Review of COVID-19 Efforts and Initiatives

This study explored initiatives through an informal search strategy to identify key initiatives at each level; however, the authors do not claim the list to be exhaustive. The authors' experience in evidence-based medicine and translational research made it easy to reach out to well-known data resources and look for COVID-19–related efforts. The study did not distinguish the type of data and information pertinent to a specific initiative; instead, it focused on resources. For different resources, different search strategies were employed. For example, primary-level studies were searched in top publisher networks by using the procedure described in Figure 1. For secondary- and tertiary-level resources, a mixed method of free searching, word of mouth, and expert recommendations were used.

Figure 1. Publisher-wise searching for published research articles on COVID-19. NEJM: New England Journal of Medicine, JAMA: Journal of the American Medical Association.



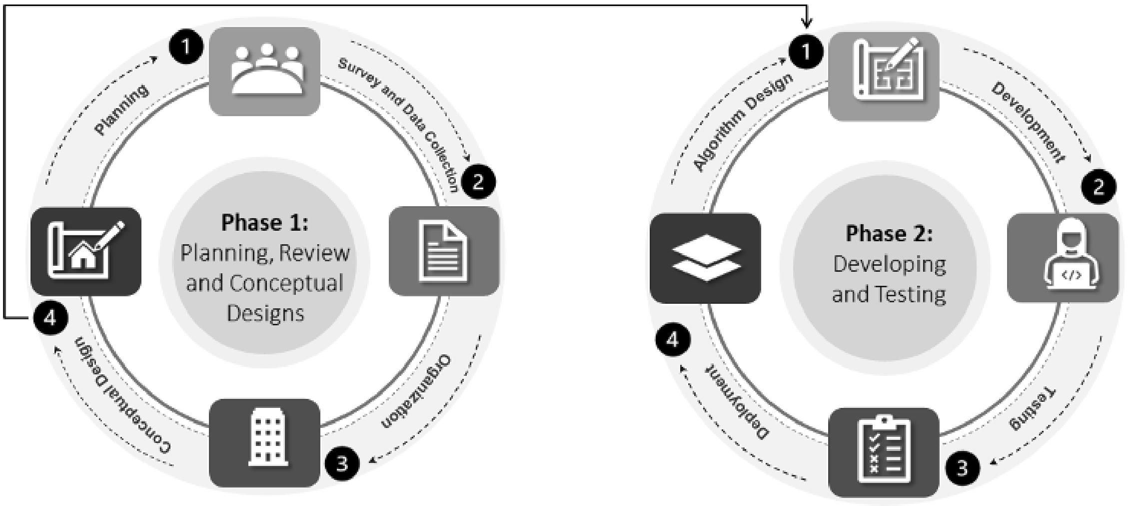
Research Design

This research reports on phase 1 of an action research project. We designed a conceptual framework for categorizing COVID-19 resources and surveyed the efforts and initiatives that contributed to creating these resources. This first phase establishes the infrastructure for locating and placing the right information at the right place. It provides a schema guideline for developing implementation models and systems.

Phase 2 involves the development of a software system and testing the viability of the proposed approach by undertaking the software development process model for each component and deploying on to an open repository for global access.

Figure 2 provides an overview of the two phases of the action research design. Planning includes brainstorming sessions, informal and formal meetings, and surveying COVID-19 efforts and initiatives to collect data. The initial data were organized to get a sense of the data elements through visuals that included bar graphs, pie charts, and histograms. Conceptual design is the design of an architecture framework for the categorization of resources. Algorithm design involves knowledge graphs, data management and querying, data-driven approaches, and user interface and user experience. Development is the actual implementation of the designed methods in the chosen programming languages. Deployment is the reflection and availability of developed methods and models to the community via open-access platforms like GitHub. As described, at any stage of phase 2, the process of phase 1 can be accessed and replicated.

Figure 2. Action research phases.



Results

COVID-19 Efforts and Initiatives

Primary Resources

The primary level includes resources that accumulate raw data or research about patients with COVID-19 and bring it into the global space. Primary-level resources were categorized into three groups: research literature, data, and clinical trials. Research literature includes published information like journal publications, raw data includes firsthand patient information like demographic and clinical information, and clinical trials includes registered trials with the objective to prove the effectiveness of a treatment like a drug or surgery. We provided a statistical overview of different resources in each group.

Research Literature

The research literature group was divided into two categories: preprint and postprint. Preprint articles are “work in progress” research that are yet to be published in a peer-reviewed journal. Postprint articles are “submitted to journal” research that are considered published after going through copyediting and typeset formatting.

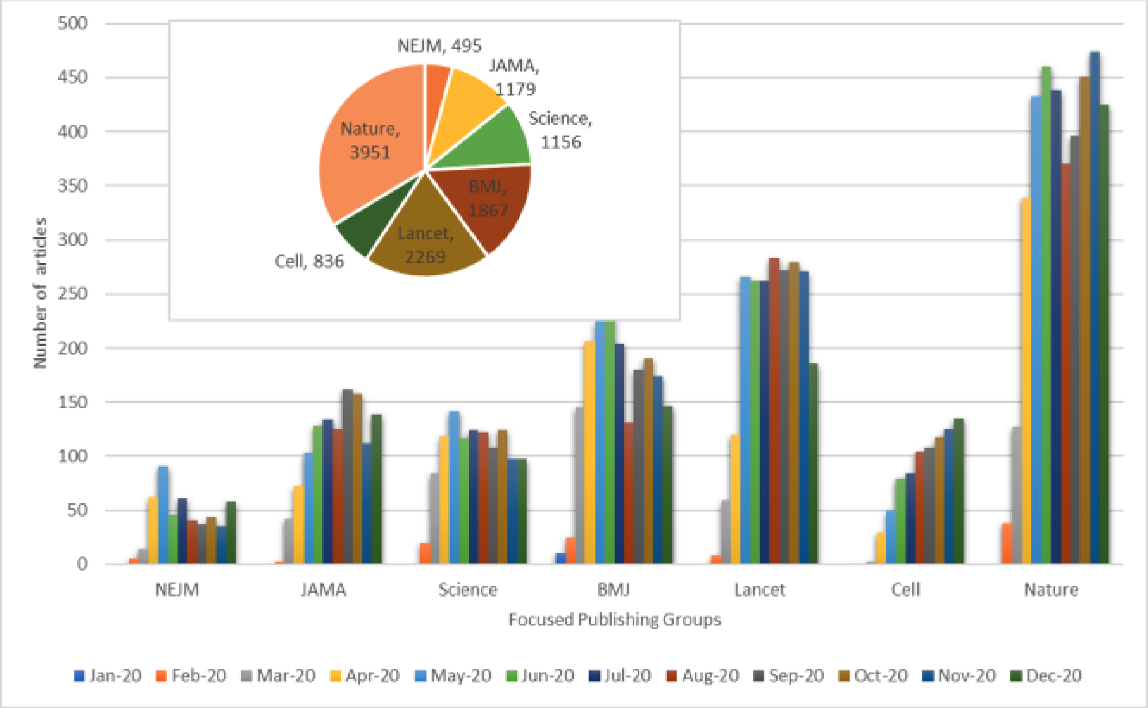
Preprint Literature

Alongside published research literature, many research articles were submitted to preprint repositories, which are unreviewed. They may eventually be published in a peer-reviewed journal. Due to unpredictable review durations, most researchers first submit an earlier version of their paper to a preprint database to disseminate their work quickly. Among popular preprints, arXiv [7] is an extensive multidisciplinary archive with 1.7 million scholarly articles in physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. It published 2842 articles on COVID-19 in 2020. Related to the medical and biology domains, two archives—medRxiv and bioRxiv [8]—are at the forefront in the collection unreviewed literature. As of 2020, medRxiv and bioRxiv published 9487 and 3058 articles, respectively. Other archives like ChemRxiv, a preprints server for chemistry and related areas, has also published a considerable number of prereviewed articles on COVID-19 (n=364).

Postprint (Published) Literature

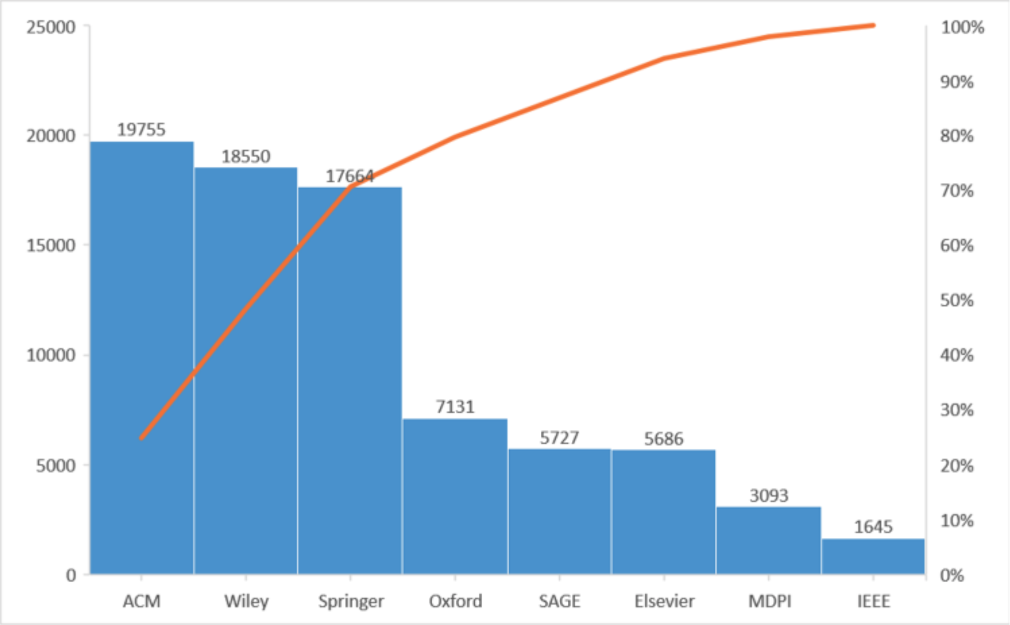
At present, a vast number of publications in the biomedical literature mention “COVID-19” in the title, the body of text, the keywords, or metadata. From January 1 to December 31, 2020, about 12,000 articles were published by prominent publishers that include Nature, Science, Cell, the New England Journal of Medicine (NEJM), the Journal of the American Medical Association (JAMA), Lancet, and BMJ. As shown in Figure 3, Nature published a total of 3951 articles, contributing about 34%, followed by Lancet (19%), in the given group of 7 publishers. It should be noted that this quantitative representation does not discuss the quality of articles. Later in this study, we will discuss COVID-19–relevant and COVID-19–specific articles. Article count increased every month initially, with the maximum number of articles appearing between May and October 2020. Starting in January 2020, there were hardly any substantive articles specific to COVID-19, although the count increased to a two-digit number in February. A maximum number was observed in October with 1364 articles. Articles continue to be published at a steady rate.

Figure 3. Monthly progression of publications on COVID-19 according to publisher. NEJM: New England Journal of Medicine, JAMA: Journal of the American Medical Association.



Many articles are being added to major libraries daily. We surveyed 8 major scientific libraries: Association for Computing Machinery (ACM), Wiley, Springer, Oxford, SAGE, Elsevier, MDPI, and IEEE. All these databases are accessible to researchers; therefore, we opted to present their statistics. We used the term "COVID-19" to keep the search simple, with increased coverage and reduced chances of missing COVID-19–related publications. As shown in Figure 4, ACM alone identified about 20,000 entries on COVID-19–related articles. Figure 4 indicates that COVID-19–related articles are published everywhere, including in IEEE Explore even though its focus is more on engineering.

Figure 4. Number of articles that appeared in major scientific libraries accessible to a researcher, presented in decreasing order from January to December 2020. The orange line displays the cumulative %, demonstrating that only three major scientific libraries (ACM, Wiley, and Springer) contributed about 70% of the articles. ACM: Association for Computing Machinery.



Patient Data Sets

Data play a crucial role in the ability to research, study, and explore population health and safety, and this is truer than ever in the context of a global pandemic. Access to data sets and associated tools that can examine that data are increasingly crucial to the research process and are particularly necessary for the worldwide response to the novel coronavirus.

To aid researchers, developers, and analysts in the struggle to fight COVID-19, different platforms offer data sets about patients with COVID-19 like the Johns Hopkins Center for Systems Science and Engineering (CSSE) [9], the Google COVID-19 Public Dataset Program [10], Kaggle [11], and GitHub [12]. Only on Kaggle are there 972 data sets of small (n=773), medium (n=168), and large (n=24) sizes. Most of these data sets are available in CSV (comma-separated value) or MS Excel (Microsoft Corp) formats. On GitHub, which is primarily a repository for hosting project code files developed in different languages, such as Python, data are also stored for the code reproducibility alongside the code files. As of June 10, 2020, 4073 public repositories have been added to GitHub matching the topic of COVID-19. Several other resources, including data.world [13], AMiner [14], and IEEEDataPort [15], provide free data sets related to COVID-19. Hundreds of regional and country-specific data sets are available through different channels; however, it is not within the scope of this study to describe them here. Rather, this study aims to present platforms that host data sets across regions and countries.

For public health safety and security, a few global-level information resources such as the WHO [16] and Centers for Disease Control and Prevention (CDC) [17] are worthy of mention as they provide important information for public awareness. The WHO is a global entity that accumulates information from its member organizations and disseminates it on its website and through other public interest channels. The WHO developed a specialized tool called the Coronavirus Disease (COVID-19) Dashboard [9], which provides a rich spectrum of metadata and analytics about COVID-19.

Whether in the United States or abroad, the CDC fights diseases that are chronic or acute, curable or preventable, and caused by human error or deliberate attack. It is a rich source of information on COVID-19 in terms of symptoms, risk factors, and guidelines on social distancing, and provides important information to different stakeholders like policymakers, travelers, businesspeople, schools, health professionals, the general public, and high-risk populations. To facilitate users' understanding of the nature of data, we describe popular data set platforms in Table 1, reporting characteristics helpful for developing statistical and machine learning models for further analysis and research. In addition to defining the platform's aim and scope, we also provide a brief explanation of the nature of the data sets hosted on that platform to help users spend their time appropriately. Some platforms offer data that can be used solely for statistical analysis and reporting purposes. In contrast, others could be used for assistance in clinical decision making related to patient diagnosis, treatment, and prognosis. In Table 2, we provide summarized information about data resources on COVID-19.

Table 1. A brief description of COVID-19 platforms that offer an open repository of data sets and their scope and usefulness.

Data set platform	Description	COVID-19 data sets/repositories, n
Kaggle [11]	Kaggle is the world's largest data science community with powerful tools and resources to achieve data science goals.	972 data sets
GitHub [12]	GitHub is primarily a host for software development and version control; however, alongside the code files, the associated data are also available for code reproducibility.	4073 repositories
data.world [13]	data.world is an open data resource hub on COVID-19 with contributions from thousands of users and organizations worldwide.	40 data sets
AMiner [14]	AMiner collects all kinds of data sets about COVID-19 with daily updates. The data are open and available for download.	438 data sets
IEEEDataPort [15]	IEEEDataPort provides free data set storage and hosts different types of data sets. It provides space to host data sets related to COVID-19. It hosts a large set of tweets data (n=174,573,543) in the English language from around the globe.	15 data sets
Google COVID-19 Public Dataset Program [10]	To make data more accessible to researchers, data scientists, and analysts, Google created a COVID-19 Public Datasets Program that hosts a repository of public data sets and is free to access and analyze.	3 data sets (JHU CSSE[a] data set, global health data from the World Bank, and OpenStreetMap data)

[a] JHU CSSE: Johns Hopkins University Center for Systems Science and Engineering.

Table 2. A brief description of COVID-19–related information resources.

Organization	Description	Key information services
World Health Organization (WHO) [18]	The WHO works with 194 member states across 6 regions and from more than 150 offices, striving to combat diseases—communicable diseases like influenza and HIV, and noncommunicable diseases like cancer and heart disease.	• A COVID-19 dashboard that provides up-to-date case information, including the number of deaths and recovery • Generates advice for public awareness to reduce the chances of being infected or spreading COVID-19 • Situation reports, released daily, provide the current COVID-19 epidemiological situation and present official case and death counts and transmission classifications. As of June 16, 2020, 148 situation reports have been released • Other services include travel advice, training and exercise, technical guidance, response funds, etc
Centers for Disease Control and Prevention (CDC) [17]	The CDC is a US-based organization that aims to provide information on health safety and security threats, both foreign and domestic.	• Information on symptoms, risk factors, and social distancing • A bank of answers to important questions and many guidance materials for various stakeholders including travelers, health care professionals, etc
EU[a] Open Data Portal [19]	The European CDC publishes a data set that contains the latest available public data on COVID-19 worldwide by screening up to 500 relevant sources every day between 6:00 and 10:00 CET.	• COVID-19 cases worldwide for download • Visualizations of cases geographically, situation dashboard, and other graphical representations • Documentation as to how the data are collected, a script of R software, and webinars
Johns Hopkins University (JHU) COVID-19 Dashboard [9]	JHU experts designed a rich and interactive COVID-19 dashboard to inform the public, help policymakers create awareness, and save lives.	• An interactive dashboard for tracking global COVID-19 cases • Animated maps that show total cases, deaths, and new cases • Critical trends on how the novel coronavirus is spreading around the globe • Worldwide mortality analysis
The World Bank COVID-19 case data [20]	The World Bank provides an array of real-time data, statistical indicators, and other types of data relevant to COVID-19, particularly on the economic and social impacts of the pandemic and the World Bank's efforts to address them.	• Global poverty estimates of the impact of COVID-19 • Health nutrition and population statistics • Understanding the COVID-19 pandemic through data on indicators and worldwide cases • Map of the World Bank's operational response to the coronavirus and relevant services
DXY [21]	DXY provides timely, accurate, and authoritative real-time reports on the COVID-19 pandemic through global mapping and knowledge.	• Global mapping of coronavirus cases • COVID-19 knowledge for the public, doctors, etc
National Institutes of Health (NIH) COVID-19 Research [22]	The NIH offers a specialized service that provides the latest research information on COVID-19.	• ACTIV (Accelerating COVID-19 Therapeutic Interventions and Vaccines) • Treatment guidelines • Grants and funding information • COVID-19 testing information

[a] EU: European Union.

Clinical Trials

Clinical trials are conducted to evaluate the effectiveness and safety of medications or medical devices by monitoring their effects on a select population. Clinical trials pass through two stages: registration and publishing. Preferably, every trial should have at least one results article, even if the results are not significant or produce negative findings; however, sometimes it may be harder to publish due to publication bias [23]. A trial can be linked to a journal article through an unstructured trial-article link (may not involve unique identifiers) or a structured trial-article link (a computable link assigned with unique identifiers such as the ClinicalTrials.gov ID or the PubMed ID) [24].

Several platforms are functional such as those that register primary clinical trials related to COVID-19 and are freely accessible and searchable. The WHO International Clinical Trials Registry Platform (ICTRP) [25] was established to facilitate a network of international clinical trial registers to ensure a single point of access, the unambiguous identification of trials, and public accessibility. From January 2020 to June 2020, a total of 3163 clinical trials have been registered in the WHO ICTRP database from 18 different sources that include ClinicalTrials.gov; Chinese Clinical Trials Registry (chiCTR); Australia New Zealand Clinical Trial Registry; Clinical Research Information Service, Republic of Korea; Clinical Trial Registry–India; EU Clinical Trials Register; German Clinical Trials Register; Iranian Registry of Clinical Trials; International Standard Randomised Controlled Trial Number Registry (ISRCTN); Japan Primary Registries Network; Lebanese

Clinical Trial Registry; Netherlands Trial Register; Pan African Clinical Trials Registry; Brazilian Clinical Trials Registry; Peruvian Clinical Trial Registry; Cuban Public Registry of Clinical Trials; Sri Lanka Clinical Trials Registry; and Thai Clinical Trials Registry.

The US-based ClinicalTrials.gov [26] is perhaps the largest database of privately and publicly funded clinical studies conducted worldwide. In this database, we found a total of 2172 studies related to COVID-19 as of June 18, 2020. The second largest clinical trials database is the chiCTR [27], a nonprofit Chinese platform that has registered 721 clinical trials. All the registries are region or country specific except the ISRCTN Registry, which initially focused on randomized controlled trials; however, but has now widened its scope to include other study types to evaluate the efficacy of human-related health interventions [28]. As of June 18, 2020, the ISRCTN database registered 70 clinical trials, out of which only 6 have been completed, 4 have been suspended, and 60 are in progress.

Figure 5 shows a progress bar of trials registered in the first 6 months of 2020. The highest number of trials (n=967) had been registered in April. The number started declining after April, which was generally seen across all databases except the chiCTR, where the decline began after February. This may be because, in China, the peak number of COVID-19 cases was observed in February. The number of trials in May and June should be higher than in April as the number of positive cases continues to rise globally.

Figure 5. Progression of the number of trials registered in major databases between January and June 2020. chiCTR: Chinese Clinical Trials Registry, WHO ICTRP: World Health Organization International Clinical Trials Registry Platform.

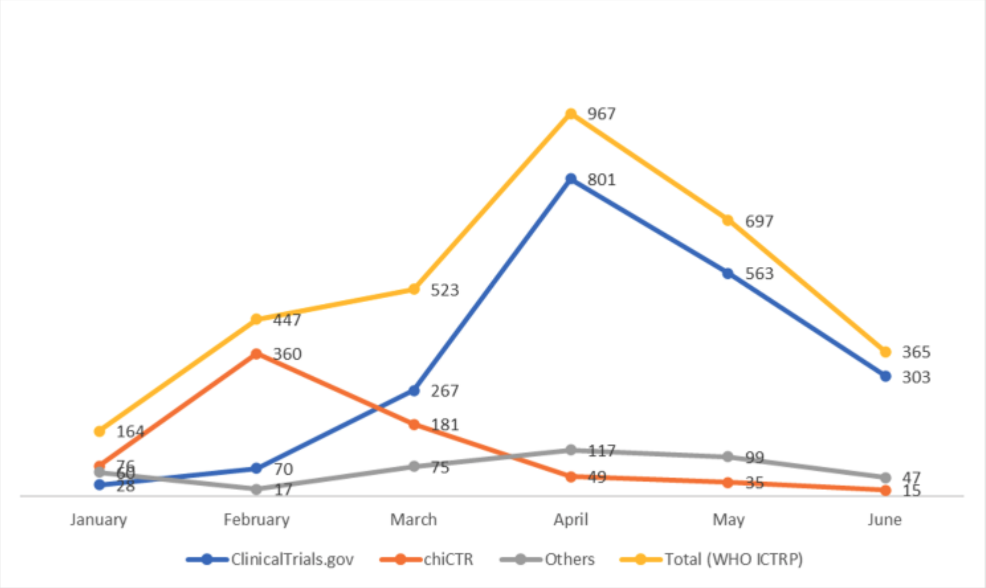
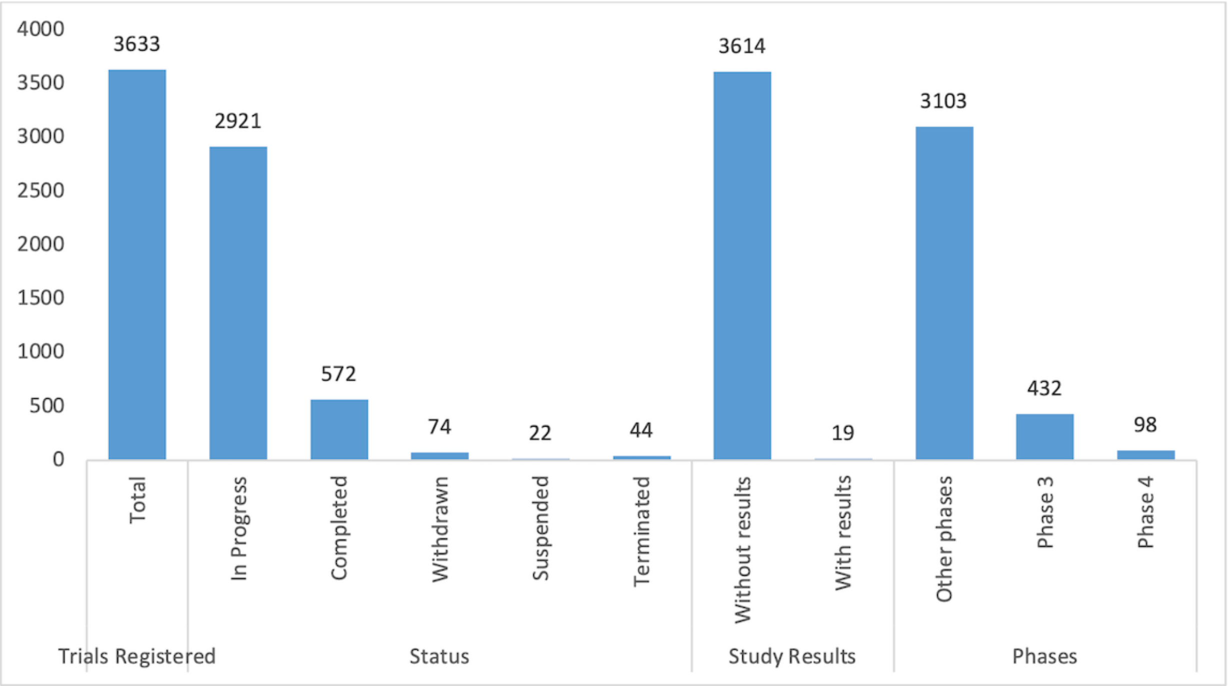


Figure 6 provides a statistical summary of ClinicalTrials.gov that includes the number of trials in 2020. The data show that only 572 trials have been completed, out of which only 19 have been reported with results. Phase IV, also called a postmarketing surveillance trial, occurs after the Food and Drug Administration has approved a drug for marketing and is the only trial that is eligible for observing drug use in public. At present, the number of phase IV trials (n=98) is far less than the phase III trials (n=432) and other trials in other phases. However, it is a reasonable number to raise the hope of obtaining significant results from watching the effects of drug on a large number of patients with COVID-19 around the world.

Figure 6. Statistical summary of clinical trials registered in the ClinicalTrials.gov database according to the trial status (eg, completed, withdrawn, suspended, terminated, or in progress), trials with results, and trials registered in different phases.



Secondary Resources

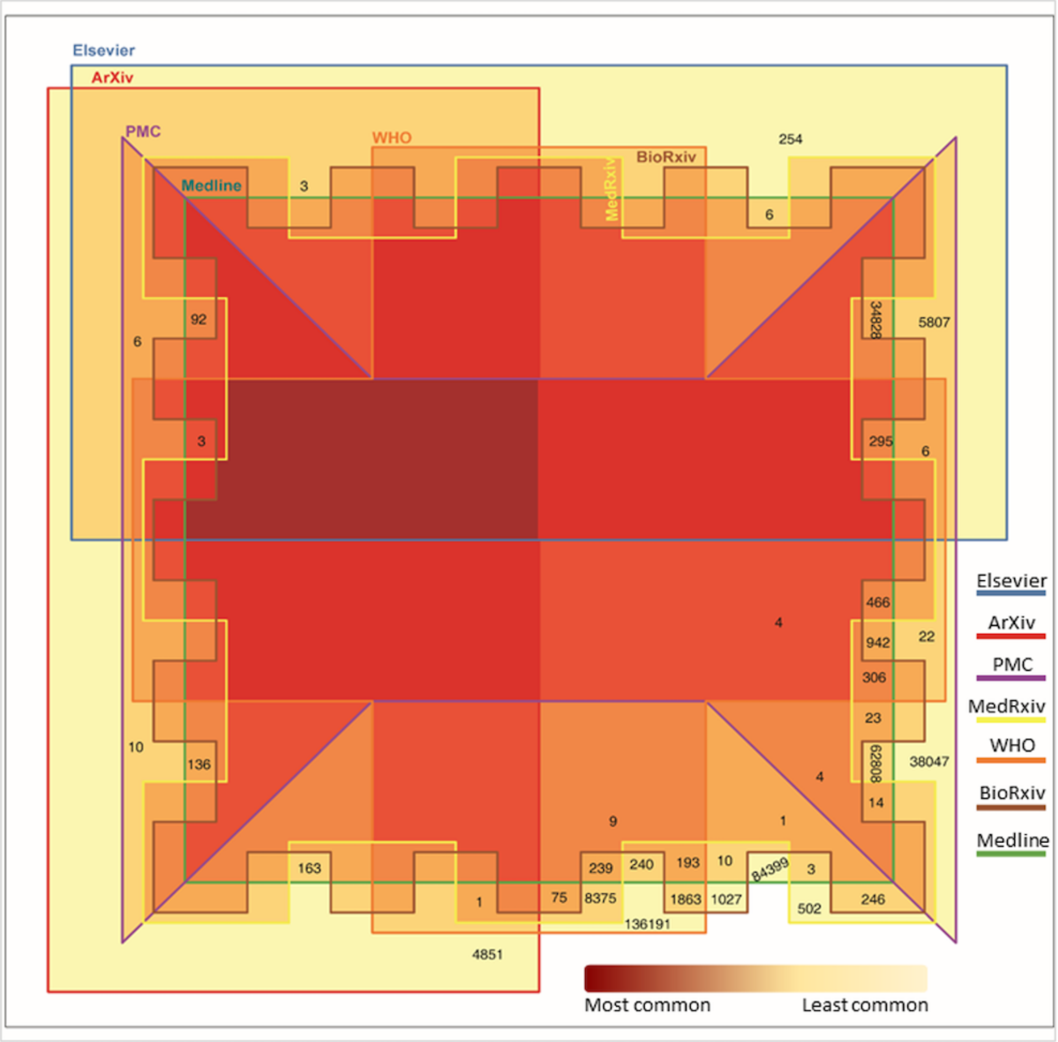
In this section, we present global-scale initiatives founded on the data acquired from primary-level sources at the secondary level. These initiatives are not discussed in any specific order. Most of these initiatives were pertinent to creating metadata on research literature articles for subsequent analysis and research.

CORD-19

CORD-19 (COVID-19 Open Research Dataset) is a free, open research data resource consisting of 130,000 scholarly articles about the novel coronavirus available for the global research community [29]. CORD-19 is updated every week with newly published research to facilitate the development of text mining and information retrieval systems, and it has been downloaded over 75,000 times in the first month of its release [29]. The articles in CORD-19 are derived from four primary-level repositories that include PubMed Central (PMC) [30], the bioRxiv and medRxiv preprint servers, and the WHO COVID-19 Database. The significant accomplishment of CORD-19 is the cleaning of metadata and machine readability of the full text. A simple deduplication logic of creating clusters for retaining similar articles, unless there is a conflict, is applied after metadata from each source is cleaned and formatted into CORD-19. After cleaning, the content is parsed from PDF-formatted papers into a JSON (JavaScript Object Notation) schema, which is simple to utilize for different text-mining tasks.

One of the CORD-19 data set features is the article's source, which represents the name of the publisher. Overall, 7 unique sources are enlisted; however, multiple sources are mentioned for articles published in more than one source. The 7 individual sources comprise articles that are either unique to them or are shared with other sources. In Figure 7, the articles shared among different resources are visualized using a Battle Venn diagram where the common area represents the number of shared articles. The intensity of color shows the high number of sources that share those articles. For instance, Medline has 84,399 articles that are not published anywhere else. However, it shares several articles with other sources; for example, with PMC, it shares 62,808 articles. No article is common in all the sources; however, 4 articles were shared among all the sources except Elsevier and arXiv. The R programming code and data files are available from the "COVID-19-Resource-Categorization" GitHub repository [31].

Figure 7. Number of articles published unique to each resource with no intersection (ie, these articles were not published in more than one resource). PMC: PubMed Central, WHO: World Health Organization.

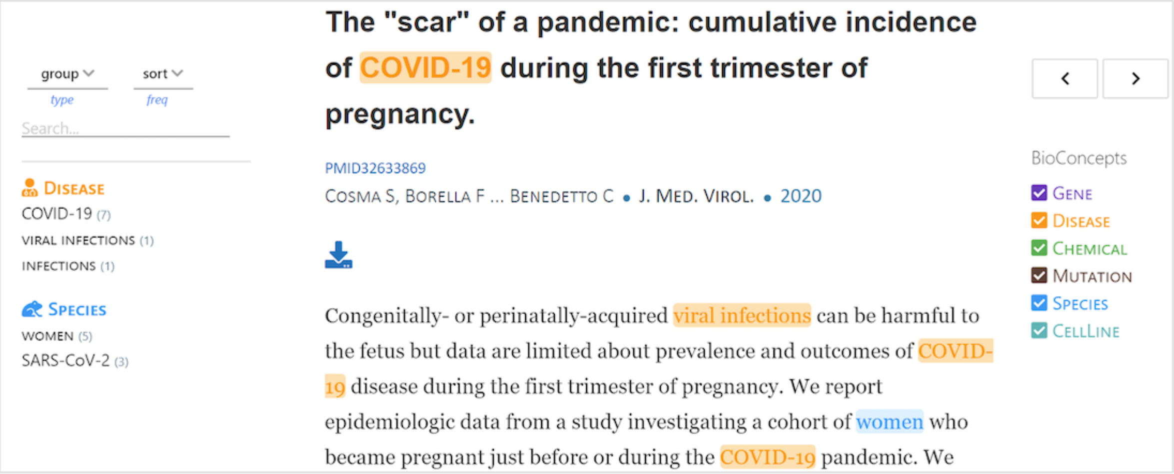


LitCovid

LitCovid is a hub of curated literature of scientific articles about COVID-19 [32-34]. The source is updated daily and has access to about 29,000 (and still growing) articles in PubMed. One unique feature of this effort is that identification of relevant articles is 35% better than the conventional keyword-based searches [34]. For improved accessibility, additional information has been added. The articles are categorized by research topics that include general information, mechanism, treatment, case report, transmission, disease diagnosis, prevention and epidemic forecasting, and geographical locations [34]. Under each topic, three additional important pieces of information are indexed—chemicals (the name of chemical products, like Remdesivir, used in different trials), journals, and countries (the host country of the research).

An associated effort extended by PubTator Central [35] derives articles from LitCovid and annotates them with 6 entity types, also called bio concepts—gene, disease, chemical, mutation, species, and cellline. The annotations of these entity types are made in different colors (Figure 8). For example, if a disease concept appears in the title or abstract, it is highlighted in yellow. Similarly, chemical names appear in green and a gene in purple. These annotations support text- and data-mining activities and are available for download in XML (Extensible Markup Language) format.

Figure 8. Annotated entity types highlighted in different colors in the title and abstract.



The "scar" of a pandemic: cumulative incidence of COVID-19 during the first trimester of pregnancy.

PMID32633869

COSMA S, BORELLA F ... BENEDETTO C • J. MED. VIROL. • 2020

Congenitally- or perinatally-acquired viral infections can be harmful to the fetus but data are limited about prevalence and outcomes of COVID-19 disease during the first trimester of pregnancy. We report epidemiologic data from a study investigating a cohort of women who became pregnant just before or during the COVID-19 pandemic. We

COVID-19 Evidence Alerts

COVID-19 Evidence Alerts is a McMaster University service that alerts users to current best evidence about COVID-19 [36]. It notifies users about the reports published in MEDLINE-based journals, which are critically appraised for scientific merit. Each appraised report is assigned a study category: diagnosis, etiology, treatment, prognosis, guideline, and clinical prediction. Reports with sufficient scientific merit are appraised as "higher-quality studies for clinical attention." Studies that provide lower-quality scientific evidence are posted as well with at least one reason for their lower-quality designation (eg, not a randomized controlled trial). There are also reports that are relevant but may not be assigned to a specific study category.

This service is beneficial for applications that involve clinical decisions. As of June 30, 2020, about 2700 reports had been posted accumulatively in all three categories; this list is growing and updated every weekday. Out of all the appraised reports, about 70 belong to the higher-quality category; this demonstrates that less than 3% of reports meet the criteria for clinical relevance and scientific merit; 34% did not meet the criteria for scientific merit, and 63% did not belong to a study category. There were two times more studies that did not belong to a study category than those that belonged to at least one study category.

CALC-19

The COVID-19 Advanced Literature Classifier (CALC-19) is a classifier of medical literature about COVID-19. The data are updated every week; as of July, the data set included more than 150,000 scholarly articles where each article is tagged by country, year, source, topic, and keywords [37]. CALC-19 extracts all of its articles from the CORD-19 data set and adds metadata like MeSH (Medical Subject Headings) keywords, authors, date of publication, and journal of publication in addition to the title and abstract. Users can search for articles of interest using different filters and tags provided by the service. The searched articles can be downloaded in Excel, CSV, Endnote, RIS, BibTex, and a new Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR) JSON format. The CALC-19 service is powered by the PICO Portal, which provides a platform to accelerate research and innovation by leveraging artificial intelligence and creating efficient systematic reviews of studies.

Evidence Aid COVID-19

Evidence Aid [38] is a collection of summaries of high-quality research studies. The collection is available in English with translations in 5 languages: French, Spanish, Portuguese, Arabic, and Chinese. This collection comprises summaries of systematic reviews to provide a quick overview of relevant and impactful information about COVID-19 concerning health conditions, outcomes, and other aspects necessary for the recovery period.

Each summary of a systematic review consists of subsections such as "Citation," "What is this?" and "What was found." The citation section includes information on the title of the systematic review, authors, publishing venue, date, and page information. The "What is this?" section summarizes information about population, experimental setup, environment, location, etc. The "What was found" section describes the findings and outcomes of the trials included in the systematic review (eg, the impact of a drug on patients with COVID-19 as positive, negative, or no effect).

Comparative Summary of Initiatives

The initiatives discussed as a part of the secondary-level resources are summarized in Table 3 to provide a comparative analysis of different aspects, such as dependent resource information, applications, language support, human validations, and download formats.

Table 3. Comparative summary of secondary-level resources based on different features that are important for researchers to expand their work on COVID-19.

Initiative or feature	Baseline resource	Direct applications	Update frequency	Multilingual support	Human expert verified	Export format
CORD-19[a]	Original research	IR and TM[b]	Weekly	N/A[c]	N/A	CSV[d]
LitCovid	PubMed	IR and TM	Daily	N/A	Yes	RIS[e], TSV[f]
COVID-19 Evidence Alerts	CORD-19	Systematic review	Every weekday	N/A	Yes	N/A
CALC-19[g]	CORD-19	IR and TM	Weekly	N/A	Users can filter records and tags to find records relevant to their work	Excel, CSV, Endnote, RIS, BibTex, HL7 FHIR[h] JSON[i]
Evidence Aid COVID-19	Systematic review	Evidence-based medicine, clinical decision support, guidelines	N/A	Yes	Yes	N/A

[a] CORD-19: COVID-19 Open Research Dataset.

[b] IR and TM: information retrieval and text mining.

[c] N/A: not applicable.

[d] CSV: comma-separated value.

[e] RIS: Research Information Systems.

[f] TSV: tab-separated value.

[g] CALC-19: COVID-19 Advanced Literature Classifier.

[h] HL7 FHIR: Health Level Seven Fast Healthcare Interoperability Resources

[i] JSON: JavaScript Object Notation.

Tertiary Resources

The tertiary-level initiatives use the outcomes of secondary resources for creating clinical guidelines, standards, and vocabularies to provide direct assistance to medical professionals and implementers of clinical decision support systems. The WHO and the CDC are the two comprehensive resources that provide general and all technical guidelines to the public and health professionals. On top of this, other resources add, refine, and customize their findings. The WHO publishes technical guidelines in 14 different categories (eg, clinical care, infection prevention and control, laboratory and diagnosis, etc) [39]. The CDC offers guidelines for various stakeholders; for health workers alone, it provides guidance in terms of 12 categories (eg, testing, clinical care, infection control, etc) [40]. Numerous initiatives derive guidance from the WHO and CDC guidelines, such as Duodecim: EBM Guidelines Coronavirus Infections [41] and COVID-19 guidelines listed in DynaMed [42].

COVID-19 Knowledge Accelerator

COVID-19 Knowledge Accelerator (COKA) [43] was first initiated in late March 2020. Led by Brian Alper, COKA is an add-on effort to an ongoing project, Evidence-Based Medicine on FHIR [44], to accelerate the processing of massive research data on COVID-19 in order to summarize and synthesize the evidence in a standard format for computable expression. It aims at resolving inefficiencies in current scientific dissemination systems in which research data are transformed into various noncomputable forms for human displays [45]. This initiative rightly identified the problem area of noncomputable communication and channelized its efforts to construct computable (structured) results directly from research publications, thus accelerating evidence synthesis.

The initiative gained momentum over a short period of time, and as of July 2020, COKA had more than 150 working meetings with more than 40 active participants from more than 25 organizations from academia, industry, government, and nonprofit organizations in 7 countries. Activities under this initiative are divided into 3-team setups: project, process, and system. Under each setup, four workgroup meetings are held every week (for a total of 12 meetings/week). Participants actively contribute to different meetings of their choice every week, and a report of the discussion is shared at the end of each day.

One of the key achievements of this initiative is the development of citation resource schemas and instances. As of July, COKA has created more than 36,000 citation resources for biomedical publications in the CORD-19 data set. Moreover, a profile resource, *EvidenceReport*, is another important outcome of this initiative, an extension of the *Composition* resource. The EvidenceReport resource provides a comprehensive report referring to one or more than one resource(s). As of July, more than 30 example reports related to COVID-10 have been generated. In addition to resource schemas, COKA made tremendous efforts in vocabulary mappings for evidence-related resources. In a short time, COKA developed a 13-step Code System Development Protocol in September 2020. HL7 FHIR is used as the underlying standard to meet interoperability needs and support the global development of terminologies for the exchange of scientific evidence [46].

ACTS COVID-19 Guidance-to-Action Collaborative

The AHRQ (Agency for Healthcare Research and Quality) Evidence-Based Care Transformation Support (ACTS) initiative of the COVID-19 Guidance-to-Action Collaborative aims to improve the development, dissemination, and use of “living” COVID-19 guidance [47]. The collaborative supports the COVID-19 “knowledge supply chain,” that is, the data-to-evidence-to-knowledge-to-guidance-to-action sequence to make the processes of guidance development, workflow integration, and knowledge supply chain more efficient and effective. Among its primary functions, the collaborative provides current solutions to urgent clinical challenges faced by health professionals, helps guidance developers in tracking COVID-19–related recommendations, fosters collaboration among implementers of COVID-19 guidance summaries, and facilitates coordination to optimize the flow of COVID-19 “evidence to action.”

The collaborative produces rapid guidance summaries that provide a comprehensive description of existing evidence and guidance from various sources such as the CDC, the WHO, and the European CDC. The guidance summaries are not clinical practice guidelines; therefore, they should not be used or interpreted as such. Instead, they can help develop local recommendations and policies. As of July, 20 guidance summaries have been produced under two categories (ie, patient care and operations). The guidance summary is structured with a question form at the top addressing major recommendations, followed by a list of evidence collected from various sources.

National Institutes of Health COVID-19 Treatment Guidelines

In collaboration with other organizations, the National Institutes of Health has developed treatment guidelines to support 297 clinicians in caring for patients with COVID-19 [48]. These guidelines are updated frequently due to the quickly evolving nature of clinical information on the new coronavirus. The guidelines’ recommendations, which are based on scientific evidence and expert opinion, possess two ratings: the strength of the recommendations indicated as a letter (A, B, or C), and the quality of the evidence indicated using a Roman numeral (I, II, or III). A panel composed of experienced representatives from 14 different organizations and societies, such as the American College of Chest and Emergency Physicians, the Food and Drug Administration, and the Society of Critical Care Medicine, has been established to develop these guidelines. The panel utilizes data from the published scientific literature on COVID-19 and the experience of its members to develop the recommendations in these guidelines.

The panel develops recommendations in clinical care areas, that is, care of critically ill patients with COVID-19, including antiviral therapy, immune-based therapy, and adjuvant therapy guidelines for special populations such as pregnant women and children. The panel’s approach to publishing the recommendation in these guidelines can be learned from the example recommendation provided about chloroquine or hydroxychloroquine in the category of antiviral therapy: “The Panel recommends against the use of high-dose chloroquine (600 mg twice daily for ten days) for the treatment of COVID-19 (AI).” This example tells us that the strength of the recommendation is “A,” which means that the statement is a strong recommendation, and the quality of evidence is “I,” which means it is supported with data from one or more randomized trials with clinical outcomes and validated laboratory endpoints.

American College of Surgeons Elective Case Triage Guidelines for Surgical Care

The American College of Surgeons (ACS) has developed recommendations for surgeons to identify which procedure should be curtailed [49]. The ACS releases newsletters to update recommendations on curtailing the performance of surgical procedures continuously 2 times a week. As of July, the ACS has published guidelines in 14 categories (eg, cancer surgery, gynecology, neurosurgery, urology, vascular surgery, etc). Most of the guidelines are provided in a descriptive form, but a few include vascular surgery and orthopedic procedures. An example of COVID-19 guidelines for the triage of vascular surgery patients is provided in Table 4, describing the meaning of each tier class (1, 2a, 2b, and 3) concerning surgery postponement.

Table 4. Examples of COVID-19 guidelines for the triage of vascular surgery patients.

Category	Condition	Tier class
AAA[a]	Ruptured or symptomatic TAAA[b] or AAA	3: Do not postpone
Peripheral aneurysm	Asymptomatic peripheral aneurysm	2a: Consider postponing
Bypass graft complications	Revascularization for high-grade restenosis of previous intervention	2b: Postpone if possible
Carotid	Asymptomatic carotid artery stenosis	1: Postpone

[a] AAA: abdominal aortic aneurysm.

[b] TAAA: thoracoabdominal aortic aneurysm.

COVID-19–SNOMED Clinical Terms

SNOMED International, a leading health care terminology organization, took steps to identify codes for different terms related to COVID-19 [50]. In the March 2020 interim release, 24 records had been added, which increased to 49 records as of August 11, 2020. A complete list of these concepts with Uniform Resource Identifiers, fully specified name, and preferred term is now available on Confluence [51]. The two most important terms, “SARS-CoV-2” (organism) and “COVID-19” (disorder) have been given the identifiers 840533007 and 840539006, respectively. A map of SNOMED Clinical Terms (SNOMED CT) to ICD-10 (International Classification of Diseases, Tenth Revision) is provided for two concepts: “COVID-19” and “Exposure to SARS-CoV-2,” as described in Table 5.

Table 5. SNOMED CT (SNOMED Clinical Terms) to ICD-10 (International Classification of Diseases, Tenth Revision) map.

Preferred term	Source SNOMED CT identifier	Target ICD-10 identifier
COVID-19	840539006	U07.1
Exposure to SARS-CoV-2	840546002	Z20.8

Summary of Tertiary Resources

This section provides an inexhaustive list of initiatives that have contributed to the creation of tertiary resources. As shown in Table 6, the initiatives are categorized into three categories: standard, guidelines, and terminologies. Information on the group or organization has been provided to support future applications to catalogue these resources in their services.

Table 6. Summary of tertiary resources related to COVID-19–based standards, guidelines, and terminologies.

Category and organization/group	Initiatives
Standards	
Health Level Seven (HL7)	HL7 is a health information exchange standard. Using its current standard—Fast Healthcare Interoperability Resources (FHIR)—different initiatives have been taken place related to COVID-19 that include COVID-19 Knowledge Accelerator (COKA) [43] and the Situational Awareness for Novel Epidemic Response (SANER) Project using HL7 FHIR to enable easier reporting for public health [52].
International Organization for Standardization (ISO)	The ISO has compiled a list of freely available standards to support global efforts in dealing with the COVID-19 crisis [53]. It also features a list of national resources developed by ISO members in different countries to support the fight against COVID-19 [54].
American National Standards Institute (ANSI)	In response to the COVID-19 pandemic, the ANSI has initiated the Standards Alliance Phase 2 (SA2) [55], which aims to reduce importation and regulatory barriers of COVID-19 testing kits and training resources on the use of medical devices and testing equipment.
Guidelines and recommendations	
National Institutes of Health (NIH)	The NIH has developed a comprehensive set of COVID-19 treatment guidelines [48].
American College of Surgeons (ACS)	The ACS Elective Case Triage Guidelines for Surgical Care [49] aims to develop recommendations to help surgeons identify which procedures should be curtailed.
World Health Organization (WHO)	The WHO is perhaps the largest source of information, guidance, and recommendations to support the fight against COVID-19. In addition to general public health recommendations, it provides technical guidance on different topics [39].
Centers for Disease Control and Prevention (CDC)	Like the WHO, the CDC offers advice, recommendations, and guidelines for different stakeholders. For health workers alone, it provides guidance on 12 categories that include testing, clinical care, infection control, etc [37].
American Gastroenterological Association (AGA)	The AGA has developed recommendations based on the systematic review and meta-analysis of 47 studies and 10,890 unique patients with gastrointestinal symptoms [56].
US Food and Drug Administration (FDA)	The FDA provides guidance related to drug development programs and the food industry impacted by COVID-19, such as donating COVID-19 plasma and facilitating the development and availability of medical device therapeutics to combat COVID-19 [57].
Vocabularies and terminologies	
SNOMED Clinical Terms (SNOMED CT)	As of August 11, 2020, SNOMED International has added 49 records related to COVID-19 [50].
International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)	The ICD-10-CM is an official coding and reporting guideline that provides important information needed to understand the usage of ICD codes in the context of COVID-19 [58]. These codes can be searched in a newly released user-friendly browser [59].
Unified Modeling Language System (UMLS)	The UMLS provides a set of COVID-19–related terms [60] mostly mapped to SNOMED CT and Medical Subject Headings.
Medical Subject Headings (MeSH)	A new MeSH Supplementary Concept Record (SCR Class 3-Disease) was added on February 13, 2020, to the 2020 MeSH Browser in response to COVID-19 [61]. The most current updates are found on the MeSH Browser [62]. Using MeSH terms, the recommended search strategy for retrieving COVID-19–related biomedical studies is “2019-nCoV OR 2019nCoV OR COVID-19 OR SARS-CoV-2 OR ((Wuhan AND coronavirus) AND 2019/12[PDAT]:2030[PDAT]).”
Logical Observation Identifiers Names and Codes (LOINC)	LOINC introduced 264 codes in response to COVID-19 [63]. These codes are distributed in 5 categories: SARS-CoV-2 lab tests (84 terms), LOINC terms for SARS-CoV-2 ask-at-order-entry questions (9 terms), convalescent plasma (2 terms), LOINC terms related to public health case reporting (63 terms), and COVID-19 or telehealth documents (106 terms).

COVID-19 Dashboard Applications

An array of dashboards has been proposed and implemented to portray the influx of data and information related to COVID-19. Here, we first discuss a set of well-known dashboards in use and provide information on different forms that include structured, unstructured, plane, and graphical.

Johns Hopkins University Interactive COVID-19 Dashboard

In response to the COVID-19 emergency, the CSSE at Johns Hopkins University (JHU), in the United States, developed an online interactive dashboard to visualize and track coronavirus cases in real time [51,64]. Starting on January 22, 2020, the dashboard data were updated manually, but with the increase in the number of cases, the manual reporting process became unsustainable and a semiautomated living data stream strategy has been used since February 1. This dashboard relies on several data sources (eg, for the identification of new cases) and seeks data from DXY China (it was initially the only data source), Twitter feeds, online news, and direct communication sent to the dashboard. The case numbers are duly confirmed with regional and local health departments, including the respective CDCs, health departments, and the WHO. This dashboard's main feature is real-time interactivity, which enables users to see cumulative cases, active cases, and other information like incidence rate on a global map. Users can click and select a region on the map and see that region's statistics. A US map is provided separately with county-wise confirmed cases and deaths. Moreover, it supports data in motion that show daily trends of cases and deaths in different regions.

WHO COVID-19 Dashboard

Like the JHU CSSE interactive dashboard, the WHO provides a live dashboard with COVID-19 case numbers and deaths where users can use their mouse to hover over a map to obtain an overview of cases across the world [16]. It provides two kinds of maps: a bubble map and a choropleth map. Alongside the maps are data tables detailing country-wise cumulative confirmed cases, deaths, as well as newly reported cases and deaths in the last 24 hours. It facilitates region-wise searches (eg, Europe, Asia, and Africa) as well.

Worldometer Coronavirus Updates

Run by an international team of developers, researchers, and volunteers, Worldometer provides global COVID-19 real-time statistics on data collected, analyzed, and validated from thousands of sources worldwide [65]. The data are claimed to be trusted and used by different governments (eg, the United Kingdom, Thailand, Pakistan, Sri Lanka, Vietnam, etc) and private organizations (eg, JHU CSSE, the BBC, the New York Times, etc). It provides country-wise statistics of new cases, deaths, recovered cases, critical cases, tests per million, and other important information. The most informative feature is the searchable and clickable country-wise data table that can be customized to include or exclude columns such as new cases, total cases, deaths, and tests per million, which are presented along with the total population of that country.

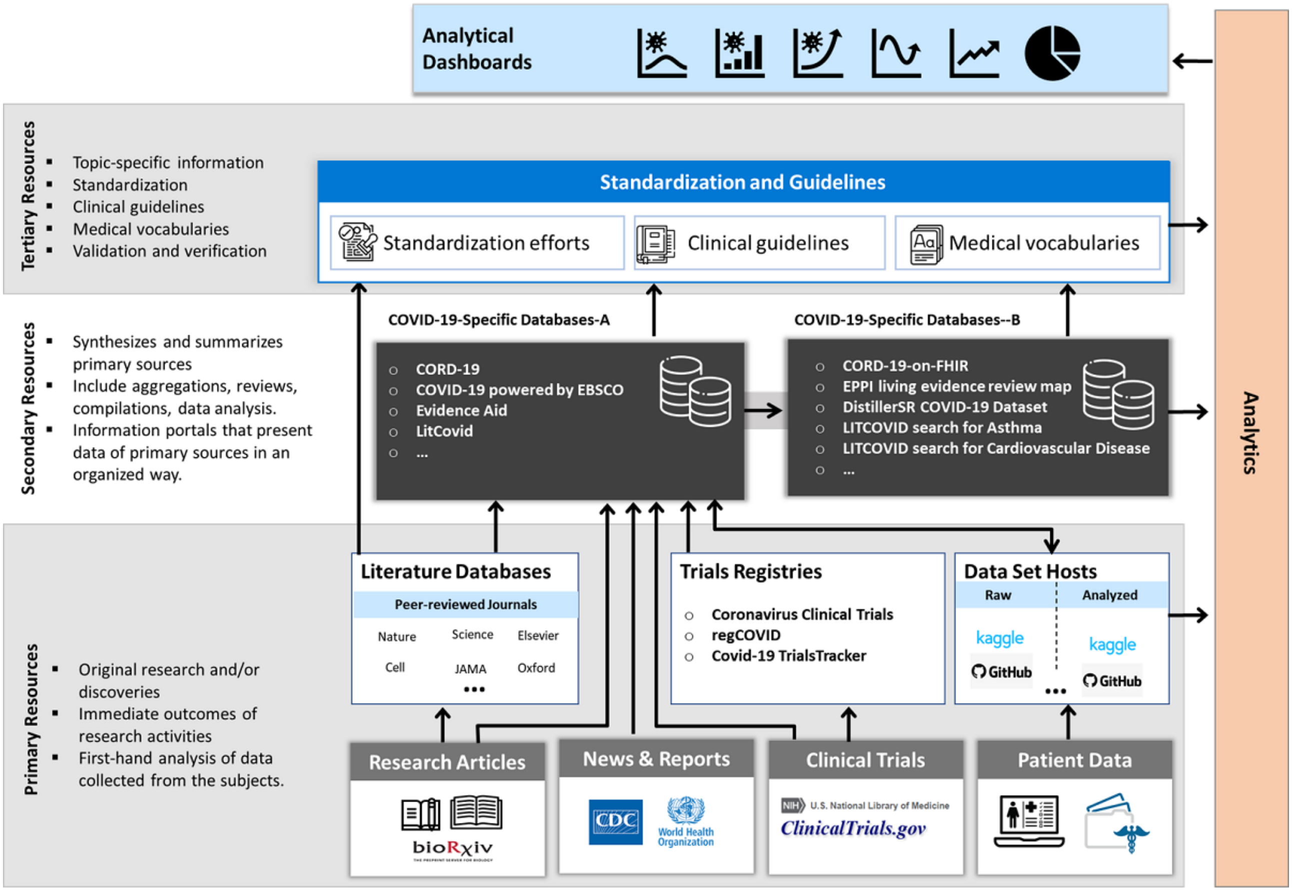
Other Efforts

Other than the dashboards mentioned above, dozens of dashboards are available for use scoped by a specific country, territory, or region. Coronaboard [66] provides COVID-19 statistics globally for the United States and three other countries (the Netherlands, France, and South Korea) in their respective languages. Almost all countries affected by the COVID-19 pandemic have put forth efforts to keep their populace informed via easy-to-use and understandable communication methods. For instance, Japan has developed its own dashboard [67], which supports the Japanese language and other languages such as English, German, French, and Arabic. Moreover, different media groups such as CNN, the BBC, and the New York Times have devised sophisticated dashboards, which rely heavily on JHU data.

A Conceptual Framework COVID-19–Related Knowledge Resources

Investigating diverse information at the three levels discussed earlier enabled us to develop an enterprise architecture framework for COVID-19–related knowledge resources. As illustrated in Figure 9, the framework comprises the categorization of efforts and initiatives, and organizes resources at the primary, secondary, and tertiary levels, demonstrating resource interconnectedness and flow of information.

Figure 9. An enterprise architecture framework for COVID-19–related efforts and initiatives involving primary, secondary, and tertiary knowledge resources. CORD-19: COVID-19 Open Research Dataset, FHIR: Fast Healthcare Interoperability Resources, EPPI: Evidence for Policy and Practice Information and Co-ordinating Centre, JAMA: Journal of the American Medical Association.



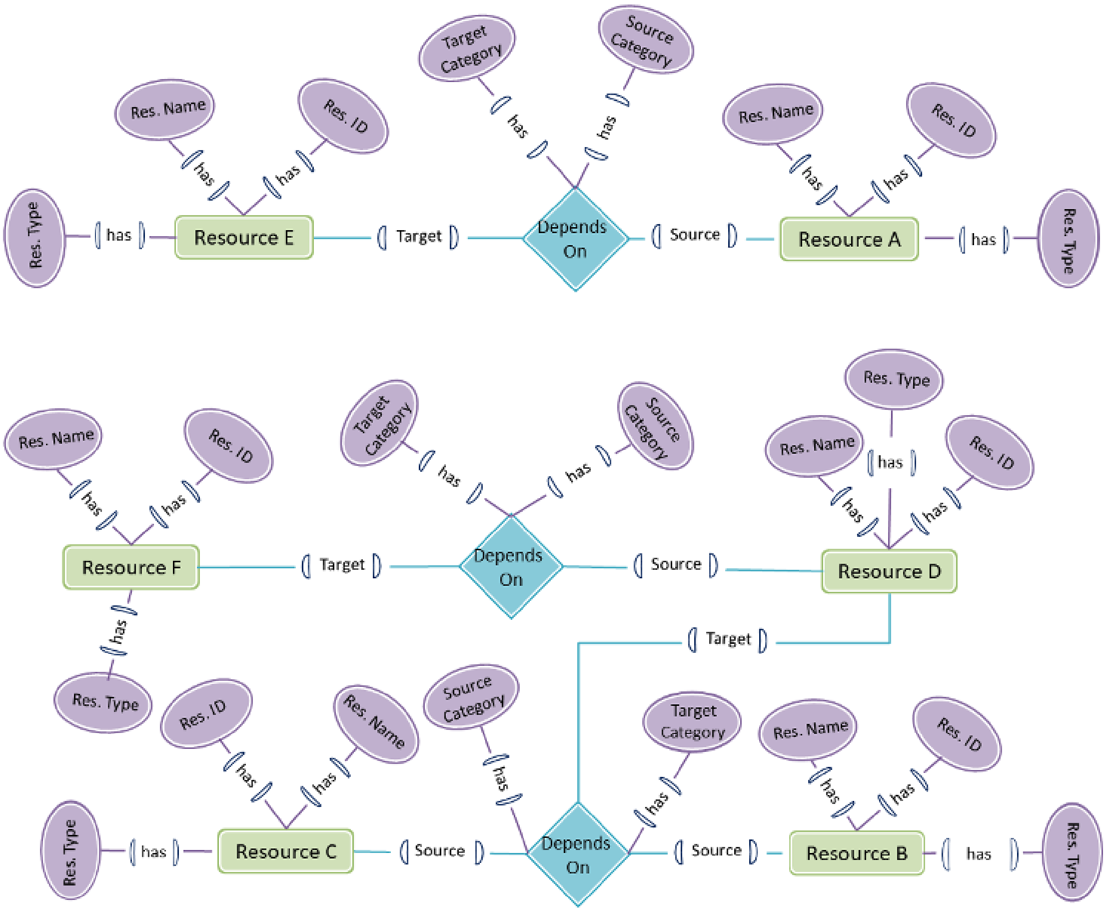
The framework's key feature is to enable tracking of a context in terms of information resources at different levels. For instance, if a user is interested in information that is an outcome of a secondary-level initiative, the dashboard will provide the required information resource(s) along with the dependent resource(s) at the primary level. The resources are then visualized in the form of dependency graphs generated automatically in response to the user query.

Figure 9 is a high-level conceptual illustration of the organization of COVID-19 resources in three logical layers connected to analytics that enable the creation of user applications such as analytical dashboards. In the following sections, we discuss the physical design of resources at different levels with relationships and properties, which demonstrates the approach's practicality.

Conceptual Modeling Using Knowledge Graphs

A robust and semantically enriched model is needed to manage the highly interlinked COVID-19 information resources and metadata. Based on the proposed framework, we created a knowledge graph design to represent interlinked resources to put data in context via linking and semantic information. An example scenario is presented in Figure 10 to display the suitability of the knowledge graph. Let us have 6 resources named A, B, C, D, E, and F, where three resources (A, B, and C) belong to primary, two resources (D and E) to secondary, and one (F) to tertiary. Let resource F be dependent on resource D, which depends on resources B and C. Similarly, let resource E be dependent on resource A. The corresponding nodes in the knowledge graph to represent the dependency among resources is shown using the relationship node *depends on*, which has two attributes, *source category* and *target category*. The source refers to the category of resources it depends on, and the target refers to the category of resource created as a result.

Figure 10. A knowledge graph representation reflecting an example scenario of 5 resources A, B, C, D, and E with their mutual dependencies.



Each resource has at least three attributes—identifiers, name, and type—where the type represents the subcategory within the primary, secondary, or tertiary category. For instance, resources whose type is data resources are research publication resources in the primary category.

Maintenance of Knowledge Graphs

The rich feature set of knowledge graphs that include classes, relationship types, and categorization support greatly help to resolve the resource maintenance challenge. Adding a new resource to the database by assigning the correct class of primary, secondary, and tertiary requires the users to input metadata about the resource. The input data may consist of dichotomous questions to support structured queries run on the underlying graph database to infer the target class for the resource, and saving a new resource must look for duplication by checking multiple parameters.

Visualization Using Knowledge Graphs

Current COVID-19 dashboards typically represent time-series visuals and geographic maps, with exceptions of dashboards that respond to the pandemic by showing clinical trials, policy- and finance-related interventions, and social distancing directives [68]. Some dashboards include contact tracing data; however, resource dependency–tracking dashboards are not seen in the literature. Novel knowledge graph visualization algorithms are emerging, customized to support resource tracking with their dependencies and metadata semantics.

Discussion

We conducted this study to classify COVID-19 resources into a three-level structure. We designed a two-phase research design approach to locate COVID-19 resources and their placement, and conceptualize the architecture in the first phase while developing and testing the contextual dashboards in the second phase. We surveyed various efforts and initiatives worldwide, provided descriptive statistics, and classified them into primary, secondary, and tertiary categories. The proposed categorization led us to design knowledge graph–based models for developing contextual dashboards. A dependency graph theory was incorporated to visualize the results of information resources and their interdependence. The proposed work enables other applications such as search engines, interactive dashboards, and tracking systems to capture contextual information of COVID-19–related resources. Using knowledge graph models for application development will add transparency to the information infrastructure, thus increasing the trust factor. Moreover, the semantics captured in subgraphs of the whole COVID-19 resource categorization knowledge graph can provide more domain information and thus significantly improve the performance of machine and deep learning models [69].

The knowledge graph has the capabilities to store a resource and associated metadata about the resource. The metadata is linked to the resources via a well-established relationship mechanism defined in the knowledge graph. These relationships provide a baseline context to the stored resources, making them more explicit and distinctive from each other. Using existing relationships in the knowledge graph could create new types of relations in the knowledge graph. Using reasoning on these extended contextual relationships could help recognize and classify the COVID-19 resources into three explicit categories. In addition, we can seek help from human experts to create an initial set of resources with explicit categorization, which can be used as a seed input to develop a machine learning model. The learned machine learning model can be used as a classification model for resource categorization, and the feature set can be used to create entities and relationships in the knowledge graph.

Moreover, knowledge graph–based representation will be a great source of assistance in data preparation for machine learning models. For instance, the source and target entities associated with each other through a relation communicate a context that is useful in word embeddings. Using machine learning over other graph databases lacks two main components: (1) the metadata related to data, which is coined as lacking a piece of contextual information; and (2) the lack of a functional module, also called the reasoning module, to interpret or transform the actual data in the presence of metadata. The knowledge graph as a source to machine learning exposes the actual data, metadata, and contextual relationships. The additional meta information and contextual relationship improve machine learning outcomes in three main ways. First, the model created from actual data associated with contextual information provides more intelligent predictions or accurate classification. Second, based on the learning patterns of machine learning, the knowledge graph reasoning capability enhances the model with additional rules and generates other contextual relationships. Third, if required (and this depends on the design of the knowledge graph), some information may be inferred by the knowledge graph during modeling and embedded to the learning phase of the machine learning method.

This study’s resource categorization can help develop applications for various purposes for combatting the novel coronavirus. First, it provides a baseline for the further categorization of resources in each level. Further classifications can be introduced in each level to make its use more specialized and customized. Second, it allows for the development of customized search engines for users to obtain results more precisely. Third, it enables specialized dashboards constructed on the information to be structured into three levels. Finally, the proposed framework is extendable to bring clinical and genomic resources together by highlighting their associations to help disease monitoring and tracking.

Acknowledgments

This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, through the Information Technology Research Center (ITRC) Support Program (IITP-2021-2017-0-01629), and in part by the Institute for Information & Communications Technology Promotion (IITP), through the Korea Government (MSIT) (2017-0-00655, IITP-2021-2020-0-01489, and NRF-2019R1A2C2090504).

Conflicts of Interest

None declared.

References

1. Kalhori SRN, Bahaadinbeigy K, Deldar K, Gholamzadeh M, Hajesmaeel-Gohari S, Ayyoubzadeh S. Digital Health Solutions to Control the COVID-19 Pandemic in Countries With High Disease Prevalence: Literature Review. J Med Internet Res 2021 Mar 10;23(3):e19473 [FREE Full text] [doi: 10.2196/19473] [Medline: 33600344]
2. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
3. Bernardo T, Sobkowich K, Forrest R, Stewart L, D'Agostino M, Perez Gutierrez E, et al. Collaborating in the Time of COVID-19: The Scope and Scale of Innovative Responses to a Global Pandemic. JMIR Public Health Surveill 2021 Feb 09;7(2):e25935 [FREE Full text] [doi: 10.2196/25935] [Medline: 33503001]
4. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
5. Alderman J. Primary, Secondary, and Tertiary Sources. University of North Florida. 2012. URL: https://digitalcommons.unf.edu/bliss/4/ [accessed 2021-05-22]
6. Elayyan M. From Books to Bytes: Navigating the Research Ecosystem. New York, NY: Center for Teaching and Learning, Columbia University; 2020 Sep 1. URL: https://academiccommons.columbia.edu/doi/10.7916/d8-t2wa-he85 [accessed 2021-05-19]
7. McKiernan G. arXiv.org: the Los Alamos National Laboratory e‐print server. International Journal on Grey Literature 2000 Sep;1(3):127-138. [doi: 10.1108/14666180010345564]
8. Sever R, Roeder T, Hindle S, Sussman L, Black K, Argentine J, et al. bioRxiv: the preprint server for biology. bioRxiv. Preprint posted online November 6, 2019 2021 [FREE Full text] [doi: 10.1101/833400]

9. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
10. About COVID-19 Public Datasets. Google Cloud Platform. 2020. URL: https://console.cloud.google.com/marketplace/details/bigquery-public-datasets/covid19-public-data-program?_ga=2.50179686.-1918167096.1591790334 [accessed 2020-06-10]
11. Search. Kaggle. URL: https://www.kaggle.com/search?q=COVID-19 [accessed 2020-06-10]
12. Search – COVID-19. GitHub. 2020. URL: https://github.com/search?q=COVID-19 [accessed 2020-06-10]
13. There are 54 covid 19 datasets available on data.world. data.world. 2020. URL: https://data.world/datasets/covid-19 [accessed 2020-06-11]
14. COVID-19 Open Datasets. China Engineering Science and Technology Knowledge Center | AMiner. 2020. URL: https://www.aminer.cn/data-covid19/ [accessed 2020-06-11]
15. COVID-19. IEEE DataPort. URL: https://ieee-dataport.org/topic-tags/covid-19 [accessed 2020-06-11]
16. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/?gclid=Cj0KCQjwiYL3BRDVARIsAF9E4GdZ0og0f2VbDJL6b6GGh9uzAKSWemLaej8CEN01Sv-_MHMu8eprTpYaAligEALw_wcB [accessed 2020-06-11]
17. Singh J. Centers for Disease Control and Prevention. Indian J Pharmacol 2004;84(12):1035-1036. [doi: 10.1097/01.phm.0000189841.18419.16]
18. About WHO. World Health Organization. 2019. URL: https://www.who.int/about [accessed 2020-06-17]
19. COVID-19 Coronavirus Data - Datasets. data.europa.eu. URL: https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data [accessed 2020-06-11]
20. Understanding the Coronavirus (COVID-19) pandemic through data. The World Bank. URL: http://datatopics.worldbank.org/universal-health-coverage/coronavirus/ [accessed 2020-06-11]
21. COVID-19 Global Pandemic Real-time Report. DXY - DX Doctor. URL: https://ncov.dxy.cn/ncovh5/view/pneumonia [accessed 2020-06-18]
22. Coronavirus (COVID-19). National Institutes of Health. 2020. URL: https://www.nih.gov/coronavirus [accessed 2020-06-18]
23. Simes RJ. Publication bias: the case for an international registry of clinical trials. J Clin Oncol 1986 Oct;4(10):1529-1541. [doi: 10.1200/JCO.1986.4.10.1529] [Medline: 3760920]
24. Huser V, Cimino JJ. Linking ClinicalTrials.gov and PubMed to track results of interventional human clinical trials. PLoS One 2013 Jul 9;8(7):e68409 [FREE Full text] [doi: 10.1371/journal.pone.0068409] [Medline: 23874614]
25. International Clinical Trials Registry Platform (ICTRP). World Health Organization. URL: https://www.who.int/ictrp/en/ [accessed 2020-06-18]
26. Search of: COVID-19 - List Results. ClinicalTrials.gov. URL: https://tinyurl.com/6c9za47r [accessed 2020-06-18]
27. Chinese Clinical Trial Registry (ChiCTR). ChiCTR.org. 2018. URL: http://www.chictr.org.cn/abouten.aspx [accessed 2020-06-19]
28. ISRCTN - Search results. ISRCTN Registry. URL: https://www.isrctn.com/search?q=COVID-19 [accessed 2020-06-18]
29. Lu Wang L, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. CORD-19: The Covid-19 Open Research Dataset. arXiv. Preprint posted online Apr 22, 2019 2021. [Medline: 32510522]
30. US National Library of Medicine - National Institutes of Health. URL: www.ncbi.nlm.nih.gov [accessed 2020-06-26]
31. Afzal M. m-afzal / COVID-19-Resource-Categorization. GitHub. URL: https://github.com/m-afzal/COVID-19-Resource-Categorization [accessed 2021-05-11]
32. 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]
33. LitCovid. 2020. URL: https://www.ncbi.nlm.nih.gov/research/coronavirus/ [accessed 2020-07-09]
34. Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. Nature 2020 Mar;579(7798):193. [doi: 10.1038/d41586-020-00694-1] [Medline: 32157233]
35. PubTator Central. URL: https://www.ncbi.nlm.nih.gov/research/pubtator/?view=docsum&query=$LitCovid [accessed 2020-07-09]
36. COVID-19 Evidence Alerts from McMaster PLUS. URL: https://plus.mcmaster.ca/COVID-19/About [accessed 2020-07-09]
37. COVID-19 Advanced Literature Classifier - PICO Portal. URL: https://covid.picoportal.net/ [accessed 2020-07-10]
38. Coronavirus (COVID-19). Evidence Aid. URL: https://evidenceaid.org/evidence/coronavirus-covid-19/ [accessed 2020-07-10]
39. Technical guidance publications. World Health Organization. 2020. URL: https://tinyurl.com/6k9zvn7s [accessed 2020-08-24]
40. Healthcare Workers: Information on COVID-19. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-nCoV/hcp/index.html [accessed 2020-08-24]
41. Duodecim: EBM Guidelines Coronavirus Infections. EBSCO COVID-19 Clinical Care. URL: https://covid-19clinicalcare.ebscomedical.com/node/59935 [accessed 2020-08-24]
42. COVID-19 (Novel Coronavirus). DynaMed. 2020. URL: https://www.dynamed.com/condition/covid-19-novel-coronavirus#GUIDELINES [accessed 2020-08-24]
43. COVID-19 Knowledge Accelerator (COKA). URL: https://www.gps.health/covid19_knowledge_accelerator.html [accessed 2020-08-20]

44. EBMonFHIR. Confluence. URL: https://confluence.hl7.org/display/CDS/EBMonFHIR [accessed 2020-08-20]
45. Alper BS, Richardson JE, Lehmann H, Subbian V. It is time for computable evidence synthesis: The COVID-19 Knowledge Accelerator initiative. J Am Med Inform Assoc 2020 Aug 01;27(8):1338-1339 [FREE Full text] [doi: 10.1093/jamia/ocaa114] [Medline: 32442263]
46. Alper B, Dehnbostel J, Afzal M, Subbian V, Soares A, Kunnamo I, COVID-19 Knowledge Accelerator (COKA) Initiative. Making science computable: Developing code systems for statistics, study design, and risk of bias. J Biomed Inform 2021 Mar;115:103685 [FREE Full text] [doi: 10.1016/j.jbi.2021.103685] [Medline: 33486066]
47. ACTS COVID-19 Guidance-to-Action Collaborative. AHRQ. URL: https://digital.ahrq.gov/covid-acts [accessed 2020-08-20]
48. COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health. URL: https://www.covid19treatmentguidelines.nih.gov/introduction/ [accessed 2020-08-25]
49. COVID-19: Elective Case Triage Guidelines for Surgical Care. American College of Surgeons. 2020 Mar 24. URL: https://www.facs.org/covid-19/clinical-guidance/elective-case [accessed 2021-05-19]
50. SNOMED CT and COVID-19. SNOMED International. URL: https://www.snomed.org/snomed-ct/covid-19 [accessed 2020-08-25]
51. SNOMED CT COVID-19 Related Content. SNOMED International / Confluence. URL: https://confluence.ihtsdotools.org/display/snomed/SNOMED+CT+COVID-19+Related+Content [accessed 2020-08-25]
52. Bazzoli F. SANER Project Using HL7 FHIR to Enable Easier Reporting for Public Health. HL7 International. 2020 Jun 25. URL: https://blog.hl7.org/saner-project-using-hl7-fhir-to-enable-easier-reporting-for-public-health-reports [accessed 2020-09-03]
53. COVID-19 response: freely available ISO standards. International Organization for Standardization. 2020 Apr 16. URL: https://www.iso.org/covid19 [accessed 2020-09-04]
54. COVID-19: National Resources. International Organization for Standardization. 2020 Nov 2. URL: https://www.iso.org/covid19-members [accessed 2020-09-04]
55. Standards Alliance Phase 2 Supports COVID-19 Response and Recovery. ANSI. 2020 Jun 19. URL: https://www.ansi.org/news_publications/news_story?menuid=7&articleid=d4868f3d-e393-4baa-94e5-9ea1e239b228 [accessed 2020-09-04]
56. Sultan S, Altayar O, Siddique SM, Davitkov P, Feuerstein JD, Lim JK, AGA Institute. Electronic address: ewilson@gastro.org. AGA Institute Rapid Review of the Gastrointestinal and Liver Manifestations of COVID-19, Meta-Analysis of International Data, and Recommendations for the Consultative Management of Patients with COVID-19. Gastroenterology 2020 Jul;159(1):320-334.e27 [FREE Full text] [doi: 10.1053/j.gastro.2020.05.001] [Medline: 32407808]
57. Coronavirus Disease 2019 (COVID-19). US Food and Drug Administration. URL: https://www.fda.gov/emergency-preparedness-and-response/counterterrorism-and-emerging-threats/coronavirus-disease-2019-covid-19 [accessed 2020-09-04]
58. ICD-10-CM Official Coding and Reporting Guidelines April 1, 2020 through September 30, 2020. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf [accessed 2020-09-04]
59. ICD-10-CM. Centers for Disease Control and Prevention. URL: https://icd10cmtool.cdc.gov/?fy=FY2020&q=covid-19 [accessed 2020-09-04]
60. Covid-19 Terms. US National Library of Medicine - National Institutes of Health. URL: https://metamap.nlm.nih.gov/Covid19Terms.shtml [accessed 2020-09-04]
61. New MeSH Supplementary Concept Record for Coronavirus Disease 2019 (COVID-19). NLM Technical Bulletin. 2020. URL: https://www.nlm.nih.gov/pubs/techbull/jf20/brief/jf20_mesh_novel_coronavirus_disease.html [accessed 2020-09-04]
62. MeSH Browser. US National Library of Medicine - National Institutes of Health. URL: https://meshb.nlm.nih.gov/record/ui?ui=C000657245 [accessed 2020-09-04]
63. SARS-CoV-2 and COVID-19 related LOINC terms. LOINC. URL: https://loinc.org/sars-cov-2-and-covid-19/ [accessed 2020-09-04]
64. COVID-19 Dashboard. Center for Systems Science and Engineering (CSSE), Johns Hopkins University. 2020. URL: https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [accessed 2020-06-18]
65. Coronavirus Live Update: Cases and Deaths from COVID-19 Virus Pandemic. Worldometer. 2020. URL: https://www.worldometers.info/coronavirus/ [accessed 2020-09-17]
66. COVID-19 Dashboard. CoronaBoard. URL: https://coronaboard.com/ [accessed 2020-09-17]
67. Japan COVID-19 Coronavirus Tracker. Reustle. URL: https://covid19japan.com/ [accessed 2020-09-17]
68. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug 07;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
69. Li R, Yin C, Yang S, Qian B, Zhang P. Marrying Medical Domain Knowledge With Deep Learning on Electronic Health Records: A Deep Visual Analytics Approach. J Med Internet Res 2020 Sep 28;22(9):e20645 [FREE Full text] [doi: 10.2196/20645] [Medline: 32985996]

Abbreviations

ACM: Association for Computing Machinery

ACS: American College of Surgeons
ACTS: AHRQ Evidence-Based Care Transformation Support
AHRQ: Agency for Healthcare Research and Quality
CALC-19: COVID-19 Advanced Literature Classifier
CDC: Centers for Disease Control and Prevention
chiCTR: Chinese Clinical Trials Registry
COKA: COVID-19 Knowledge Accelerator
CORD-19: COVID-19 Open Research Dataset
CSSE: Center for Systems Science and Engineering
CSV: comma-separated value
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level Seven
ICD-10: International Classification of Diseases, Tenth Revision
ICTRP: WHO International Clinical Trials Registry Platform
ISRCTN: International Standard Randomised Controlled Trial Number Registry
JAMA: Journal of the American Medical Association
JHU: Johns Hopkins University
JSON: JavaScript Object Notation
MeSH: Medical Subject Headings
NEJM: New England Journal of Medicine
PMC: PubMed Central
SNOMED CT: SNOMED Clinical Terms
WHO: World Health Organization
XML: Extensible Markup Language

Edited by C Basch; submitted 18.04.21; peer-reviewed by HA Khattak, D Hussain; comments to author 06.05.21; revised version received 12.05.21; accepted 12.05.21; published 01.06.21.

Please cite as:
Afzal M, Hussain M, Hussain J, Bang J, Lee S
COVID-19 Knowledge Resource Categorization and Tracking: Conceptual Framework Study
J Med Internet Res 2021;23(6):e29730
URL: https://www.jmir.org/2021/6/e29730
doi:10.2196/29730
PMID:33999833

Original Paper

Quality and Adoption of COVID-19 Tracing Apps and Recommendations for Development: Systematic Interdisciplinary Review of European Apps

Leonie Kahnbach[1,2*], MSc; Dirk Lehr[1*], Prof Dr; Jessica Brandenburger[3], MSc; Tim Mallwitz[3], MSc; Sophie Jent[4], MSc; Sandy Hannibal[1], BSc; Burkhardt Funk[5], Prof Dr; Monique Janneck[3], Prof Dr

[1]Department of Health Psychology and Applied Biological Psychology, Leuphana University of Lüneburg, Lüneburg, Germany
[2]Competencies for Digitally-Enhanced Individualized Practice Project, Leuphana University of Lüneburg, Lüneburg, Germany
[3]Institute for Interactive Systems, Technische Hochschule Lübeck, Lübeck, Germany
[4]Department for Electrical Engineering and Computer Science, Technische Hochschule Lübeck, Lübeck, Germany
[5]Institute of Information Systems, Leuphana University of Lüneburg, Lüneburg, Germany
*these authors contributed equally

Corresponding Author:
Dirk Lehr, Prof Dr
Department of Health Psychology and Applied Biological Psychology
Leuphana University of Lüneburg
Universitätsallee 1
C1.120
Lüneburg, 21335
Germany
Phone: 49 4131 677 2720
Email: lehr@leuphana.de

Abstract

Background: Simulation study results suggest that COVID-19 contact tracing apps have the potential to achieve pandemic control. Concordantly, high app adoption rates were a stipulated prerequisite for success. Early studies on potential adoption were encouraging. Several factors predicting adoption rates were investigated, especially pertaining to user characteristics. Since then, several countries have released COVID-19 contact tracing apps.

Objective: This study's primary aim is to investigate the quality characteristics of national European COVID-19 contact tracing apps, thereby shifting attention from user to app characteristics. The secondary aim is to investigate associations between app quality and adoption. Finally, app features contributing to higher app quality were identified.

Methods: Eligible COVID-19 contact tracing apps were those released by national health authorities of European Union member states, former member states, and countries of the European Free Trade Association, all countries with comparable legal standards concerning personal data protection and app use voluntariness. The Mobile App Rating Scale was used to assess app quality. An interdisciplinary team, consisting of two health and two human–computer interaction scientists, independently conducted Mobile App Rating Scale ratings. To investigate associations between app quality and adoption rates and infection rates, Bayesian linear regression analyses were conducted.

Results: We discovered 21 national COVID-19 contact tracing apps, all demonstrating high quality overall and high-level functionality, aesthetics, and information quality. However, the average app adoption rate of 22.9% (SD 12.5%) was below the level recommended by simulation studies. Lower levels of engagement-oriented app design were detected, with substantial variations between apps. By regression analyses, the best-case adoption rate was calculated by assuming apps achieve the highest ratings. The mean best-case adoption rates for engagement and overall app quality were 39.5% and 43.6%, respectively. Higher adoption rates were associated with lower cumulative infection rates. Overall, we identified 5 feature categories (symptom assessment and monitoring, regularly updated information, individualization, tracing, and communication) and 14 individual features that contributed to higher app quality. These 14 features were a symptom checker, a symptom diary, statistics on COVID-19, app use, public health instructions and restrictions, information of burden on health care system, assigning personal data, regional updates, control over tracing activity, contact diary, venue check-in, chats, helplines, and app-sharing capacity.

Conclusions: European national health authorities have generally released high quality COVID-19 contact tracing apps, with regard to functionality, aesthetics, and information quality. However, the app's engagement-oriented design generally was of lower quality, even though regression analyses results identify engagement as a promising optimization target to increase adoption rates. Associations between higher app adoption and lower infection rates are consistent with simulation study results, albeit acknowledging that app use might be part of a broader set of protective attitudes and behaviors for self and others. Various features were identified that could guide further engagement-enhancing app development.

(J Med Internet Res 2021;23(6):e27989) doi:10.2196/27989

KEYWORDS

COVID-19; contact tracing; app-based tracing; Mobile App Rating Scale; user engagement; human–computer interaction

Introduction

COVID-19 Contact Tracing Apps

For the first time, digital technology could play a key role in fighting a global health crisis. Particularly, COVID-19 contact tracing apps have been advocated as a way to achieve pandemic control, avoid or leave lockdowns [1], and return to normalcy [2]. Contact tracing and other nonpharmaceutical measures—like testing, case isolation, quarantining, hygiene, decontamination, and physical distancing—are especially important in situations where vaccinations and effective treatment are not widely available. Even in light of achievements in developing the first available vaccines [3], several countries will have been locked down twice or even three times (eg, Germany and Austria both as of December 2020) before pandemic control is expected to be reached through vaccination dissemination.

The potential advantage of contact tracing apps consists in shared characteristics of digital technology and the SARS-CoV-2 virus: speed and spread. SARS-CoV-2 is being transmitted between people at a high speed. Hinch and colleagues [4] have assumed a generation time for virus transmission of just 6 days and an epidemic doubling time of 3-3.5 days, resulting in a large proportion of the population potentially becoming infected within short periods of time in the absence of any effective intervention. For this reason, modeling study results have continuously suggested that manual contact tracing might be too slow, largely due to personnel limitations, and only feasible in locations with low incidence rates [1,2,5]. Briefly, manual contact tracing is a process by which some index person with confirmed infection reaches out to contact health authority personnel to provide information about other people with whom they have come in contact with and whom they, thereby, might have infected. The health authorities then trace those people to inform them about their possible infection and the necessity that they seek testing or quarantine themselves [5].

Major advantages of digital technology are its large scalability and its potential to speed up the tracing process. Key features of contact tracing apps for COVID-19 consist of informing others immediately of having been in contact with a person who is infected, thus speeding up isolation, testing, and quarantining. Moreover, contact tracing apps with symptom checkers can also reduce testing delay (ie, the time duration between symptom onset and subsequent testing) by immediately referring people to testing facilities [5] and reduce the time of uncertainty by providing test results more quickly. Moreover, digital contact tracing might be more reliable than manual tracing, as it is not affected by recall bias [2], especially of casual contacts, and allows for anonymous contacts to be traced [5]. Finally, once released, contact tracing apps may reach the majority of the population within a short period of time. It is estimated that 3.6 billion people worldwide already had access to a mobile device in 2020 [6]. In Europe alone, 540 million people (72% of the population) are currently in possession of a functional smartphone [6].

App Adoption and Effectiveness

Acceptance by the potential users is the major challenge for COVID-19 contact tracing apps reaching their potential. Simulation studies concordantly emphasize the importance of high rates of app adoption in the population. In a best-case scenario, Xia and Lee [2] found that 90%-95% of the population must use a contact tracing app to stop the spread of COVID-19 and allow normalcy without physical distancing. Based on data for the United Kingdom, Hinch and colleagues [4] found that the pandemic could effectively be suppressed if 80% of smartphone users, or 56% of the population, use a contact tracing app. Kretzschmar and colleagues [5] used data from the Netherlands and investigated different scenarios including adoption rates. In a best-case scenario in which the app adoption rate is 80%, almost 80% of forward transmission could be prevented. Interestingly, even with an adoption rate of 20%, contact tracing apps were more effective than manual tracing, and just 40% of the population needed to use the app to control the pandemic. Yasaka and colleagues [7] found the best results for an adoption rate of 75%, but even a rate 25% could provide some suppression of the infection curve. Similarly, Moreno López and colleagues [8] reported that a 30% adoption rate of tracing apps could be sufficient to reduce the pandemic to a manageable level if the dynamics of infection are moderate. Braithwaite and colleagues [9] presumed a quadratic relationship between a population's app adoption rate (ie, 80% adoption rate) and associated reductions in transmission (ie, 64% of contacts notified). However, assuming an adoption rate of 53%, Kucharski and colleagues [10] suggested that manual tracing is more effective than digital tracing. It should be noted that no empirical data have yet been published, with all the aforementioned numbers merely the result of modeling studies whose parameter settings might be debated (eg, assumptions made by Kucharski and colleagues [10] about the capacity of manual tracing in high incidence situations might be too optimistic). In a recent combined simulation and observational study, Wymant and colleagues [11] found that every 1% increase in app adoption lead to a decrease of 0.8% up to 2.3% in infections. To summarize, these studies suggest that high adoption rates are vital for contact tracing apps to play a key role in fighting COVID-19 but that "lower numbers of app users [would] also have a positive effect" [4].

Initial studies on the intention to download and use a contact tracing app were encouraging. In a large multinational study in Western countries (France, Germany, Italy, the United Kingdom, and the United States), the intention to use a contact tracing app ranged from 68% to 75% [12]. In a Belgian study, 49% of participants intended to use a contact tracing app [13]. In a large study in Ireland, 83% claimed that they would either definitely or probably install a contact tracing app on their smartphone [14]. Additionally, in a Dutch sample, app adoption rates varied between 59% and 66%, depending on the app's features [15].

Numerous factors have been investigated as potential predictors of app adoption. They have included potential users' age, gender, comorbid illness, smartphone use, trust in the government, opt-in installation [12], belonging to a higher risk group, general trust in others, privacy concerns [16], the presence of serious health conditions, level of education, fear of infection, expected general adoption rate in the society [15], attitudes toward protecting family members and friends, feeling responsible to the community, COVID-19–related worry [14], perceived benefits, cues to use the app in the media, and self-efficacy toward properly using the app [13].

Quality Characteristics

Although such personal characteristics, individual expectations, and societal variables are important, quality-related characteristics of contact tracing apps might also play an important role in their acceptance and adoption rates. According to Stoyanov and colleagues [17], a health app's quality characteristics can be subcategorized into those pertaining to engagement, functionality, aesthetics, and information quality. These four dimensions of COVID-19–related app quality were investigated in an early review published by Davalbhakta and colleagues [18], focusing on India, the United States, and the United Kingdom. Most apps provided information on COVID-19 or included symptom checkers. In general, the quality of COVID-19 apps was above average, and apps were designed to achieve higher scores in functionality, while features to make apps engaging and important to the user were not as highly considered. Interestingly, adoption rate (measured through the number of app downloads as a proxy for adoption) was not correlated with app quality. For Europe, Davalbhakta and colleagues [18] claimed that apps generally focus on providing high quality information from credible sources but often lack creative and interactive methods to provide this information. At the time of their review, only two COVID-19 contact tracing apps from national health authorities had been launched in the European Union. Since then, almost all European countries have developed and launched COVID-19 contact tracing apps. To the best of our knowledge, their quality characteristics and actual adoption rates have not been yet systematically investigated.

The primary aim of this study is, therefore, to investigate the quality characteristics of European national COVID-19 contact tracing apps, in terms of their engagement, functionality, aesthetics, and information quality. A secondary aim is to examine quality characteristics and their associations with app adoption rates. Likewise, we will investigate adoption rates and associated numbers of confirmed COVID-19 cases as a measure of pandemic control. Furthermore, we also will analyze features embedded within COVID-19 contact tracing apps. Finally, based on our results and the identification of apps with the highest quality scores, we will provide recommendations for optimizing contact tracing apps and increasing adoption rates.

Methods

Search Strategy

Eligible countries were member states of the European Union, former members (the United Kingdom), and countries of the European Free Trade Association (Switzerland, Iceland, Norway, and Liechtenstein). These countries share similar standards of voluntariness of app use, data protection, and privacy regulations, namely the European General Data Protection Regulation (GDPR) or regulations that are equivalent to these standards. Further inclusion criteria were a tracing app launched by a national health authority, voluntary app use, unrestricted access to the app, English version available, and app launch at least 4 weeks before the review was conducted. To identify all relevant national COVID-19 contact tracing apps, an English Google search was performed using the keywords "corona tracing app," "corona app," and "covid app" combined with the name of each country. For the United Kingdom, we considered England, Scotland, Wales, and Northern Ireland separately. For some countries, no national tracing apps were identified directly; therefore, the search was repeated in each country's main language using translation software. The apps were then downloaded through the Google Play Store or Apple App Store.

Interdisciplinary Quality Rating

Quality characteristics were assessed with the Mobile App Rating Scale (MARS) [17] using the German version [19]. The MARS quality section consists of four subscales: engagement, functionality, aesthetics, and information quality. These scales collectively contain 19 questions that reviewers are asked to respond to using a 5-point response scale (1: inadequate; 2: poor; 3: acceptable; 4: good; 5: excellent). One item directed at the level of evidence of treatments (item 19) was excluded. For each scale, the mean score was calculated, ranging from 1 to 5. The MARS has been shown to exhibit very good internal consistency (α=.90) and interrater reliability (intraclass correlation coefficient [ICC] 0.79) [17].

Four reviewers were used in this study: two researchers with a background in e–mental health and two researchers from the human–computer interaction field. All experts rated the apps independently of each other, after each rater finished the rater training proposed by Messner and colleagues [19] and practiced ratings on 3 apps for depression. The scores were combined and ICCs calculated for each app. In cases of ICCs below 0.75, discrepancies in ratings were discussed until consensus about the rating was reached, as proposed by Messner and colleagues [19]. For the ratings, two devices operating with iOS and two with Android were used.

Adoption Rates

All app developers were asked, via email, to provide the numbers of downloads. In cases of no reply, a reminder was sent after 1 week. If answers from developers could not be attained personally, the number of downloads was extracted from published statistics by national health authorities, app developers, or other publicly available sources. We calculated the percentual of the download count within the population in each country to obtain the overall adoption rate. The total download numbers, numbers of inhabitants, and data of retrieval are available in Multimedia Appendix 1.

Infection Rates

For each country, the number of infected people were obtained from publicly available data from the COVID-19 Dashboard by the Center for Systems and Engineering at Johns Hopkins University [20]. The time span from app release to the date of measuring adoption rate was considered. Numbers are reported in Multimedia Appendix 1.

Statistical Analysis

A mixed methods approach was used for this review. For the MARS total score and each subscale, the mean score was calculated and reported, ranging from 1 to 5. ICCs were calculated using a two-way mixed effects average measures model with absolute agreement [17]. To explore associations between contact tracing app quality characteristics and adoption rates, scatterplots and trend lines were presented together with univariate ordinary least squares (OLS) and Bayesian regression analysis [21]. Due to the expected low sample size of European contact tracing apps, any significance testing and relying on *P* values are highly likely to miss existing associations (type II error). Therefore, we focused on effect sizes and reported all associations exceeding a medium effect size, as recommended by Cohen [22].

Additionally, we calculated adoption rates for best-case scenarios (best-case adoption rate [BCAR]), assuming that app quality characteristics were optimized (five points on the MARS). The same analyses were conducted to investigate associations between adoption rates and measures of pandemic control.

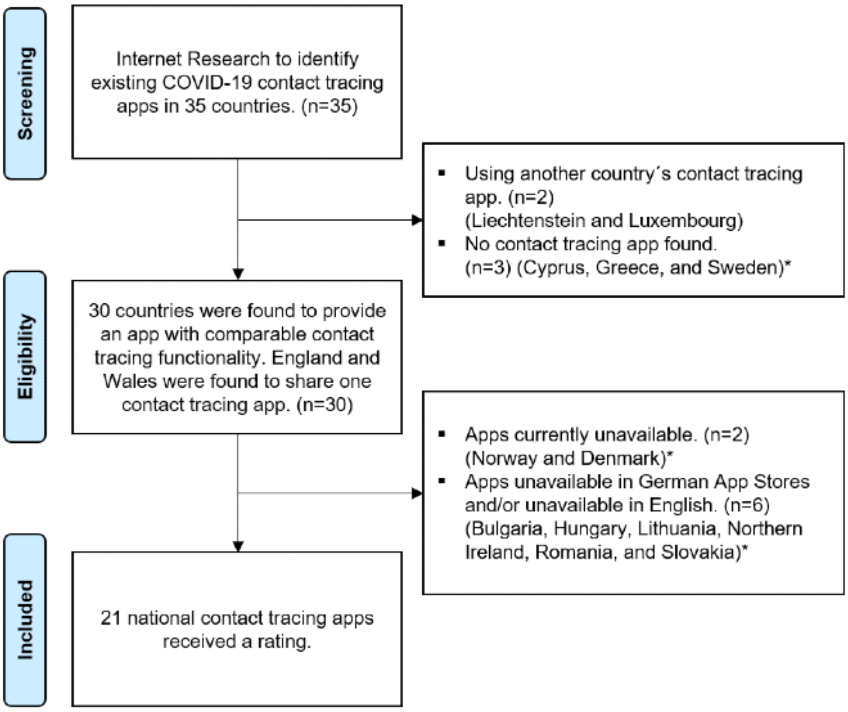
As part of our qualitative analysis, all raters were instructed to identify features of the tracing apps that, in their opinion, might increase user engagement and adoption rates. Features were identified by all raters and then discussed and classified by all authors. As a result, all apps were analyzed with respect to the features they offered. Additionally, apps with high scores on the various MARS items were presented as examples of best practice. Finally, MARS ratings, identified features, and best practice guided recommendations for optimizing acceptance and adoption of COVID-19 contact tracing apps are proposed.

Results

Search

We identified 21 contact tracing apps that we then included in our quality review. For a detailed description of the search process, see Figure 1.

Figure 1. App selection process flowchart.



*Screening was finished 1st of November 2020; Apps / English versions may be available by now.

Interdisciplinary Quality Rating

For 17 apps, ICCs between 0.75 and 0.89 were detected, indicating good interrater reliability, while excellent interrater reliability (ICC>0.90) was observed for 2 apps. Due to low ICCs, ratings for 2 apps needed to be discussed, resulting in moderate ICCs of 0.70 (France) and 0.65 (Poland) [23].

In general, the average MARS total score was 3.97, indicating above average app quality, with mean scores ranging from 3.53 to 4.42 (see Table 1 and Figure 2). The Irish app was rated best (MARS ∅=4.42), almost 1 SD above the app with the second highest rating (Finland; MARS ∅=4.24). With regard to app quality characteristics, functionality, aesthetics, and information quality scored over 4 points. High scores for information quality and functionality were especially associated with low variance, indicating small differences between the apps (Table 1). The scores for engagement were lower, on average, and the differences between apps larger. Although high scores for functionality reflect a ceiling effect and scores for aesthetics and information quality a close-to-ceiling effect, lower scores for engagement indicate opportunities for improvement.

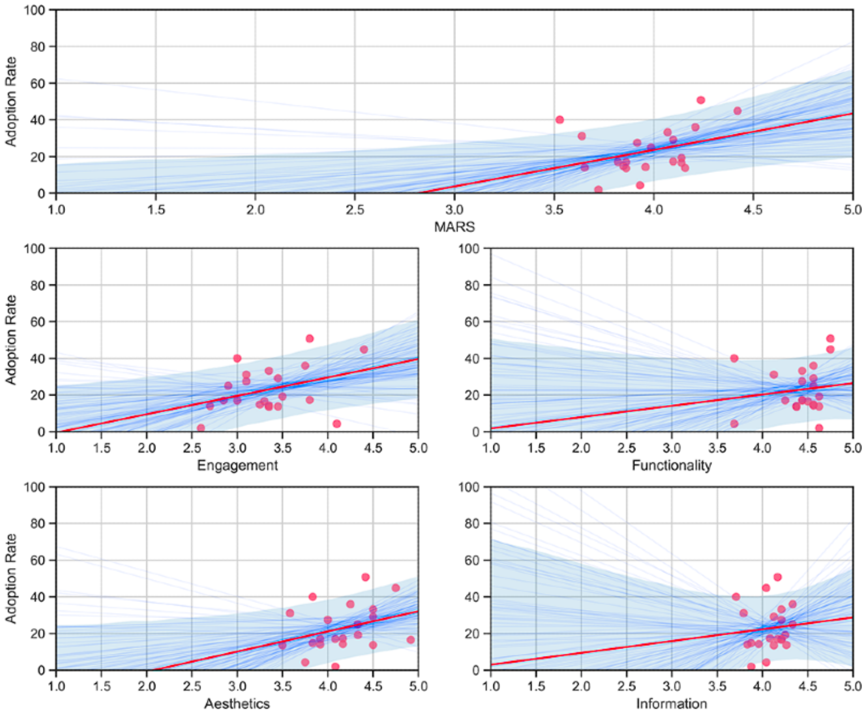
Table 1. MARS rating scores on all apps (N=21).

Country	App name	MARS[a] total	Engagement	Functionality	Aesthetics	Information	Downloads, n	Adoption rate (%)
All, mean (SD)	N/A[b]	3.97 (0.22)	3.34 (0.45)	4.43 (0.28)	4.16 (0.37)	4.09 (0.18)	97,097,051	22.92 (12.51)
Austria	Stopp Corona App	3.85	3.25	4.56	3.83	3.88	1,312,063	14.81
Belgium	Coronalert	4.14	3.50	4.63	4.33	4.25	2,200,000	19.20
Croatia	Stop COVID-19	3.72	2.60	4.63	4.08	3.88	78,534	1.93
Czech Republic	eRouška	3.86	3.35	4.38	3.50	4.13	1,443,691	13.56
England/Wales	NHS COVID-19	4.21	3.75	4.56	4.25	4.33	20,361,253	35.97
Estonia	HOIA	3.86	3.00	4.44	3.92	4.17	224,833	17.31
Finland	Koronavilkku	4.24	3.80	4.75	4.42	4.17	2,800,000	50.74
France	Tous Anti-COVID	4.10	3.80	4.44	4.17	4.08	11,600,000	17.31
Germany	Corona-Warn-App	4.10	3.45	4.56	4.50	4.13	24,200,000	29.15
Iceland	Rakning C-19	3.53	3.00	3.69	3.83	3.71	142,796	40.00
Ireland	COVID Tracker Ireland	4.42	4.40	4.75	4.75	4.04	2,200,000	44.86
Italy	Immuni	4.14	3.30	4.50	4.92	4.21	10,000,000	16.57
Latvia	Apturi COVID	4.16	3.45	4.63	4.50	4.26	263,848	13.74
Malta	CovidAlert Malta	3.82	2.85	4.25	4.08	4.21	84,210	17.06
Netherlands	Corona Melder NL	3.99	2.90	4.56	4.33	4.33	4,321,443	25.01
Poland	STOP COVID ProteGO Safe	3.93	4.10	3.69	3.75	4.04	1,637,927	4.31
Portugal	STAYAWAY COVID	3.92	3.10	4.44	4.00	4.21	2,822,522	27.47
Scotland	Protect Scotland	3.64	3.10	4.13	3.58	3.79	1,700,000	31.12
Slovenia	#OstaniZdrav	3.96	3.35	4.56	4.17	3.96	297,000	14.27
Spain	Radar COVID	3.65	2.70	4.38	3.92	3.83	6,571,600	14.00
Switzerland	SwissCovid	4.07	3.35	4.44	4.50	4.21	2,835,331	33.18

[a] MARS: Mobile App Rating Scale.

[b] N/A: not applicable.

Figure 2. Bayesian regression with trend lines and 80% CIs (blue shaded area) for app characteristics and adoption of COVID-19 contact tracing apps. MARS: Mobile App Rating Scale.



Adoption Rate

European COVID-19 contact tracing apps had an average population adoption rate of 23%, with high variation between countries, indicated by an SD of 12.5%. Taking different population sizes into account, the overall adoption rate was estimated at 22%. Adoption varied between the highest rates for Finland (51%), Ireland (45%), and Iceland (40%), and the lowest for the Czech Republic (14%). Poland (4%) and Croatia (2%) were considered low adoption outliers. Data on app release are summarized in Multimedia Appendix 1.

App Quality Characteristics and Adoption Rates

The MARS total score and two dimensions of quality characteristics exceeded the minimum threshold for an intermediate strength correlation: general app quality (MARS total score; R=0.35), engagement (R=0.35), and aesthetics (R=0.31). To calculate the BCAR, we conducted a univariate Bayesian linear regression using 20,000 Markov chain Monte Carlo draws for each app characteristic. Results are presented in Table 2 and Figure 2. This analysis yielded one BCAR distribution per app characteristic (Figure 3), resulting in a mean BCAR of 44% for apps with maximum general app quality, 40% for apps designed in the most engaging way, and 32% for apps with maximum scores for aesthetic design. Among app quality characteristics with high average scores—functionality and information quality—no associations exceeded the minimum threshold.

Table 2. Regression analysis app characteristics correlated with moderate-effect sizes for adoption of COVID-19 tracing apps (results for Bayesian regression are based on 20,000 draws).

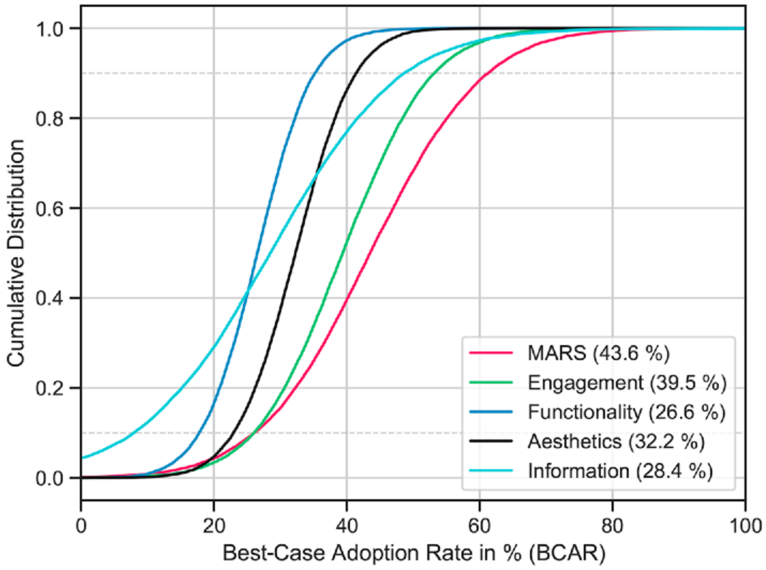
Predictor	R	Intercept, mean (SD)	*t* test (*df*)	Slope, mean (SD)	*t* test (*df*)	BCAR[a,b] 10% percentile	BCAR	BCAR 90% percentile
MARS[c] total score	0.349	–55.68 (50.43)	1.10 (19)	19.82 (12.69)	1.56 (19)	25.28	43.56	61.21
Quality characteristics (dimensions)								
Engagement	0.354	–10.53 (21.07)	0.50 (19)	10.02 (6.25)	1.60 (19)	25.77	39.54	53.69
Functionality	0.139	–4.58 (45.83)	0.10 (19)	6.21 (10.34)	0.60 (19)	17.71	26.56	34.78
Aesthetics	0.314	–22.36 (33.61)	0.67 (19)	10.90 (8.04)	1.36 (19)	22.96	32.17	41.18
Information	0.084	0.33 (68.30)	0.00 (19)	5.53 (16.71)	0.33 (19)	8.14	28.37	48.16
Specific quality characteristics (items)								
Interest (item 2)	0.405	3.48 (11.07)	0.31 (19)	5.98 (3.29)	1.82 (19)	25.37	33.32	41.17
Interactivity (item 4)	0.300	–7.31 (23.87)	0.31 (19)	9.51 (7.45)	1.28 (19)	22.87	40.18	57.43
Target group (item 5)	0.488	–37.72 (26.19)	1.44 (19)	13.99 (6.02)	2.32 (19)	26.12	32.21	38.20
Graphics (item 11)	0.324	–36.04 (40.44)	0.89 (19)	13.00 (8.89)	1.46 (19)	22.41	28.86	35.22
Accuracy of app description (item 13)	0.424	–105.55 (66.47)	1.59 (19)	26.77 (13.84)	1.93 (19)	23.29	28.30	33.27
Quantity of information (item 16)	0.413	–10.72 (38.72)	0.28 (19)	7.83 (8.99)	0.87 (19)	19.41	28.30	37.04

[a] BCAR: best-case adoption rate.

[b] BCARs are assuming maximum MARS scores.

[c] MARS: Mobile App Rating Scale.

Figure 3. Cumulative distribution of the BCAR; numbers in brackets indicate the mean BCAR for specific app characteristics. BCAR: best-case adoption rate; MARS: Mobile App Rating Scale.

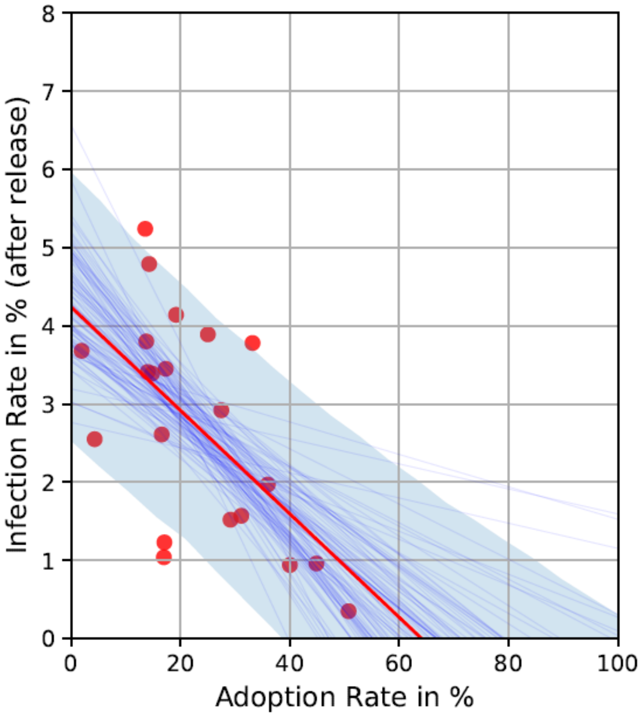


For 18 specific quality characteristics, six MARS items were correlated with higher adoption rates with at least a medium level effect (Table 2). Higher ratings for interesting use facilitating frequent engagement correlated with higher adoption (item 2; R=0.41) and an above average BCAR of 33%. More interactive apps—allowing user input, providing feedback, and containing prompts (item 4)—were directly correlated with app adoption (R=0.30; BCAR 40%). Likewise, appropriate app design for the target group (item 5) was correlated with the app adoption rate (item 5; R=0.49, BCAR 32%). These three specific app characteristics are part of the engagement dimension. Item 11 considers the quality of graphics and visual design of buttons, icons, menus, and content as part of the aesthetic dimension. Higher scores were correlated with higher adoption rates (R=0.32; BCAR 29%). The accuracy of the app description in app stores (item 13) and the quantity of information (item 16) represent facets of the information quality dimension. A highly accurate description of the components and functions of the app was correlated with higher adoption rates (R=0.42; BCAR 28%). Likewise, comprehensive and concise information with links to more information and resources was correlated positively with adoption (R=0.41; BCAR 28%). The regression model for the target group was the only one with significant effects.

App Adoption and Infection Rates

Next, the associations between app adoption rates and infection rates due to COVID-19 were investigated (see Figure 4). Results for Bayesian regression based on 20,000 draws suggested an inverse correlation between app adoption and infection rates after app release: the higher the adoption rate, the lower the rate of infections (r=–0.606; *P*=.004). Based on numbers at the time of assessing adoption (mean 22.92) and infection (mean 3.05) rates, the results indicated that a 1-point increase in adoption is associated with a 0.07 percentage points decrease in infection rate, representing a change in infection rate of 2.30%. According to the OLS regression (y = 4.23 + –0.07 * x), a zero rate of infections (trend line intersects x-axis) is reached if the adoption rate equals 60.43%. The SD of the intercept was 0.57 (t_{19}=7.36), and for the slope, it was 0.02 (t_{19}=2.99). Excluding low adopting outliers, the effects were even more pronounced for infection rates (r=–0.656; *P*=.002).

Figure 4. Bayesian regression with trend lines and 80% CI for app adoption of COVID-19 contact tracing apps and infection rate after app release.



Features of COVID-19 Contact Tracing Apps

On qualitative analysis, various features of the COVID-19 contact tracing apps were identified that might contribute to higher adoption rates. This section gives an overview of features implemented by presenting best practice examples. We focus on features that contributed to higher ratings in each MARS dimension.

Engagement

Most features aimed to increase users' engagement with the app. Table 3 provides an overview of features believed to enhance engagement beyond the core functionality of contact tracing. Apps with high scores on engagement—like those in Ireland, Poland, Finland, France, and England and Wales (see Table 1)—provided best practice examples.

Table 3. Features of COVID-19 contact tracing apps contributing to engagement.

Categories and features[a]	Countries																				
	AT[b]	BE[c]	CH[d]	CZ[e]	DE[f]	EE[g]	EN[h]	ES[i]	FI[j]	FR[k]	HR[l]	IS[m]	IE[n]	IT[o]	LV[p]	MT[q]	NL[r]	PL[s]	PT[t]	SC[u]	SI[v]
Symptoms																					
Symptom checker[w]	✓[x]	—[y]	—	—	—	—	✓	—	✓[z]	—	—	—	—	—	—	—	—	✓	—	✓[z]	—
Symptom diary[aa]	—	—	—	—	—	—	—	—	—	—	—	—	✓	—	—	—	—	✓	—	—	—
Regularly updated information provided within the app																					
Statistics on COVID-19[ab]	—	✓	✓	✓	—	—	—	—	✓[z]	✓	—	—	✓	—	✓	✓[z]	—	—	—	—	—
App use[ac]	—	—	✓	✓	—	—	—	—	—	✓	—	—	✓	—	—	—	—	—	—	✓	—
Public health instructions/restrictions[ad]	—	—	—	—	—	—	—	—	✓[z]	✓	—	✓[z]	—	—	—	✓[z]	—	✓[z]	—	—	—
Burden on health care system[ae]	—	✓	—	✓	—	—	—	—	✓[z]	✓	—	—	✓	—	—	—	—	—	—	—	—
Individualization																					
Assigning personal data[af]	—	—	—	—	—	—	✓	—	—	—	—	✓	✓	✓	✓	—	—	✓	—	—	—
Updates on assigned region[ag]	—	—	—	—	—	—	✓	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Tracing																					
Control over tracing activity[ah]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	—	✓	✓	✓	✓	—	✓	✓	✓	✓
Contact diary[ai]	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	✓	—	—	—
Venue check-in[aj]	—	—	—	—	—	—	✓	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Communication																					
Chat[ak]	—	—	—	—	—	—	—	—	—	—	—	✓	—	—	—	—	—	—	—	—	—
COVID-19–related helplines[al]	—	—	✓	✓	—	—	—	—	—	—	—	✓	—	—	—	—	✓	✓	—	—	—
Sharing[am]	✓	✓	✓	✓	✓	—	—	—	—	✓	—	—	✓	✓	✓	—	✓	—	✓	✓	✓

[a] The specified functions represent the status as of December 5, 2020. App development is dynamic and additional features may have been released in the meantime (eg, a contact diary and daily statistics have been added in the German app).

[b] AT: Austria.

[c] BE: Belgium.

[d] CH: Switzerland.

[e] CZ: Czech Republic.

[f] DE: Germany.

[g] EE: Estonia.

[h] EN: England/Wales.

[i] ES: Spain.

[j] FI: Finland.

[k] FR: France.

[l] HR: Croatia.

[m] IS: Iceland.

[n] IE: Ireland.

[o] IT: Italy.

[p] LV: Latvia.

[q] MT: Malta.

[r] NL: Netherlands.

[s] PL: Poland.

[t] PT: Portugal.

[u] SC: Scotland.

[v] SI: Slovenia.

[w] Symptom checker: symptom checker is used only if required.

[x] Included in this app.

[y] Not included in this app.

[z] Indicates that features are displayed on an external website.

[aa] Symptom diary: symptoms are monitored daily and symptom history is saved.

[ab] Statistics on COVID-19: numerical information regarding infections, deaths, tested, etc.

[ac] App use: numbers of app downloads, transferred positive COVID-19 tests, warnings from the app, etc.

[ad] Public health instructions/restrictions: app links to current information on the pandemic, for example, on how to protect oneself from the virus or which restrictions are currently in place.

[ae] Burden on health care system: numerical information on available intensive care units, numbers of COVID-19 cases hospitalized, etc.

[af] Assigning personal data: potential to provide personal data (telephone number, nickname, postal code).

[ag] Updates on assigned region: updates on the status of assigned region of residence.

[ah] Control over tracing activity: potential to disable tracing in app settings.

[ai] Contact diary: users can document their contacts.

[aj] Venue check-in: using a Quick Response code, users check into venues.

[ak] Chat: app provides a chat function.

[al] COVID-19–related helplines: app provides telephone numbers to seek help (medical).

[am] Sharing: app has a (prominently displayed) button so the app can be shared.

Interest

Several apps provided a *symptom checker.* They included a series of COVID-19–specific questions and provided individual feedback and recommendations. Typically, users could assess the occurrence and intensity of symptoms and receive recommendations for appropriate action (eg, scheduling an appointment at a test center). Users were free to check symptoms as often as they wished. Some apps addressed the benefits of checking symptoms by providing a link to an external website. Accordingly, users had to leave the app, and those apps could not make use of any data on the external website. However, most symptom checkers saved symptomatic data and included a history of symptoms within the app. In this case, a *symptom diary* was coded. The symptom diary enabled users to monitor symptoms that were displayed using a numeric and text-based design. Graphic displays to present the course of symptoms were not observed. Noteworthy is that the checking feature implemented in the Polish app was a combined *risk group checker* and symptom checker. Therefore, users must indicate age and any pre-existing illness (eg, obesity along with current symptoms), and feedback probably was provided by taking all this information into account. A separate risk group checker informing users about, for example, any increased risk of severe COVID-19 symptoms, was not identified in any app. Besides the symptom diary, a *contact diary* was observed in 1 app (Poland). Users can take written notes on each contact, which helps them to remember contacts and give more detailed information in case of infection. Other modes of contact diary input or output—like speech, picture, or location-based notes—were not found. All these checking and monitoring features lead to higher scores for engagement, as they provide added value and invite users to interact with the app more often.

Features providing new and relevant information contributed to higher scores in engagement as well. First, several apps informed the users about current numbers of newly infected people and the number of infected people who had died (see Table 3). Second, some apps provided information on public health measures, like current restrictions during a lockdown. Third, users receive information on the capacity of and burden on the health care system (eg, current number of hospitalized patients or occupied beds in intensive care units, as done with the French app). Fourth, some have features informing users about the number of downloads, thereby informing them about the size of the app-using community and motivating them to share the app with others. Fifth, a few apps additionally informed users about the protective behaviors of other users (eg, number of shared positive test results; see Swiss app) or about others' self-protective behaviors, like the number of users that conducted a symptom check that day (see Irish app). Such features were rated positively, as they can strengthen a sense of belonging to a community of people who take the pandemic seriously and care for one another. However, it should be noted that apps provided this information differently: sometimes directly included in the app, while others provided a link to an external website, thereby, interrupting users' engagement with the app.

Although paper-and-pencil lists were used in several countries to document visitors (eg, museums) or customers (eg, restaurants), the *venue check-in* (see English and Welsh app) offered a digital solution as part of their tracing app by scanning the QR (Quick Response) codes of pertinent venues. As all visitors are registered with a time stamp, it is possible to exchange information on infections quickly and reliably.

Customization

Several features were identified that contributed to higher customization ratings. First, with regard to visual customization, the Portuguese app offered users the opportunity to switch between different themes: dark mode, light mode, and automatic. In this way, users can choose the mode that appeals most to them, as opposed to most other apps, which were designed using a one-size-fits-all strategy. Second, there was one example of a customized welcome message (Polish app); with this, users could voluntarily assign themselves a nickname to receive a personalized welcome while using the app. Third, some apps provide regional customization, like the English and Welsh app, so users can receive news about the status of COVID-19 in their own region of residence. To facilitate the identification of COVID-19 hot spots, the Italian app asked users to indicate their region of residence. Fourth, users can customize the channel of support by voluntarily providing their phone number as a callback option (Irish and Latvian apps). Fifth, most apps offer to customize tracing functionality, meaning that users can pause contact tracing. Most of these customization options were voluntary. However, to be able to use the app, it was sometimes mandatory for users to indicate their postal code (England and Wales) or provide their telephone number (Iceland).

Interactivity

High interactivity facilitates interactions between users or considers the interaction between the user and the app. First, interactive symptom diaries and checkers were rated higher than noninteractive versions of these features (eg, the app links to a symptom checker on an external website). A high level of interactivity also is achieved if symptoms can be checked within the app and feedback provided based on the results, for example, with recommendations regarding self-isolation and taking a test (English and Welsh app) or friendly, encouraging feedback when no symptoms are reported (Irish app). Likewise, the interactive risk group checker provided individual feedback. Lower interactivity was deemed present in apps that merely included links to external websites to assess potential symptoms of COVID-19. Second, features that allow users to share the app with their family and friends in an easy and comfortable way contributed to higher ratings for interactivity (eg, the Dutch app). Third, 1 app provided users the option of interacting directly with health officials via chat (Iceland).

Target Group

Adequate consideration of the target group relates to intended users' sociodemographic characteristics, preferences, and needs. First, all reviewed apps were rated to be designed for the average user and the whole population, irrespective of user age, gender, or socioeconomic background. Apps received higher scores when the language used was easy to understand. Within most apps' illustrations, the icons and images supported textual information and contributed to better understanding. Likewise, almost all the apps were designed according to current trends and standards. Noteworthy is that illustrations within the English and Welsh app were designed to represent a diverse society with people from different cultures. Second, the target group can also be characterized by their needs and expectations (eg, the need for predictability, users feeling well prepared when they receive an alert, or using the app to share one's own positive test result with others). The Dutch app provides a good example for a preview on how the app will alert the user in the event of a detected encounter with a person who has tested positive. To do so, this app uses explanatory text paired with graphic support.

Functionality, Aesthetics, and Information

In general, the functionality, aesthetics, and information quality of all the apps were rated highly, with minimal variance between the ratings. Therefore, only a few features were identified that could make a difference and potentially increase users' adoption of the app. According to their MARS ratings, the rater team identified best practice examples of app features for each dimension: the Finish and Irish apps provide good examples for functionality, the Italian for aesthetics, and the Dutch and English/Welsh apps for information.

High scores for functionality were achieved by most of the apps, as they performed very well. Their structure was easy to understand because content areas were often clearly separated by card-based layouts and marked with headings. Additionally, all the apps were easy to use. In terms of navigation, some differences appeared. For example, tab navigation with icons and text in the footer and metanavigation integrated into the header was rated highly, as this configuration allows easy orientation (ie, Irish and Portuguese apps).

Most apps were rated highly with regard to aesthetics, with the graphic design of most corresponding to design trends of the years 2020 and 2021 (based on research by creative platforms, creative experts, and agencies), thereby making them attractive for most people. All apps used sans serif fonts and almost all preferred illustrations over photos. Most apps and almost all illustrations appeared in flat design, sometimes in combination with light background gradients or background circles. The character designs often appeared on monochrome backgrounds (eg, Belgium, Germany, and Slovenia) in analogue color schemes (eg, Italy and Finland) or with complementary color schemes (eg, Germany and the Netherlands), presumably to focus on the content or the message of the illustrations, rather than to illustrate the app's functionality. Disproportionate body parts embodied the character design of some apps (eg, Italy), for example, with small heads on large bodies with large limbs, while detailed facial features were largely avoided. The modern line art style was used especially for icons in most apps (eg, Finland's navigation tabs). Meanwhile, with the Estonian app, this style was used for the entire app's appearance. The use of curved lines was an exception only observed with the Latvian app. Buttons often appeared haptic, with drop shadows compared to illustrations, making it easier for users to identify elements of the interface that they can interact with.

Apps with the highest aesthetics ratings used different font styles with high contrast and strong, saturated colors for components of the interface design. Visual information groups were formed using the law of proximity, in the form of cards, spaces between components, and white space in general. Based on our assessment of the apps, visual appeal was mostly characterized by flawless graphics and a uniform and professional design. Small animations bring graphics to life (eg, Italy), stimulating positive feelings in users.

High ratings for information quality were achieved by high-level accuracy in app descriptions in app stores, reflecting the app's actual functionality. Likewise, the apps' goals were well described in the stores or within the apps. Some differences were observed in the extent to which illustrations and visualizations (tables, images, graphics) were used to deliver information. Generally, apps with more illustrations and visualizations were rated higher. Credibility of information was rated high, as all apps were published by the health authorities in their country of use. Noteworthy is the Dutch app, a good example of providing concise and helpful information on how to use the app. For example, users can easily understand how the app will warn them if they ever come in contact with a person testing positive for COVID-19.

Discussion

General Findings

This study investigated quality characteristics among European COVID-19 contact tracing apps and explored these characteristics' associations with user adoption rates. In general, scores for functionality, aesthetics, and information quality were high, and differences between the apps were minor. However, the quality of features aiming to increase users' engagement with the app was lower, on average, with substantial differences between the apps that reflected large differences in adoption rates. Alongside quality ratings, we identified engagement-friendly features implemented in tracing apps that contribute to the higher quality of those apps, may guide further development, and could increase the acceptance and adoption of contact tracing apps in the general population. These engagement-friendly features are presented within nine recommendations for an engagement-oriented design for COVID-19 contact tracing apps.

Findings in Context

Comparing the level of quality observed in our study to the earlier review by Davalbhakta and colleagues [18], there is increased quality of the COVID-19 contact tracing apps, with an overall MARS rating 0.52 points higher, a 13% increase. Especially with regard to information quality (+0.83) and aesthetics (+0.68), apps achieved higher ratings, while scores for functionality were almost unchanged. This overall improvement is probably because the apps assessed in this study were in place later in 2020, meaning that developers simply had more time to improve features like aesthetics and engagement beyond tracing core functionality. Differences in information quality might be explained by all the apps in this study having been released by the countries' own national health authorities. Interestingly, the relative order in tracing apps' quality—with functionality scoring highest, followed by aesthetics and information quality, and engagement lowest—found both in our study and the study conducted by Davalbhakta and colleagues [18] has also been observed in apps for other health conditions, like rheumatism [24], heart failure [25], and posttraumatic stress disorder [26], and for health behaviors, like physical activity [27], and might indicate a general underemphasize of engagement-friendly designs for health-targeted apps. Overall, adoption rates were substantially lower than anticipated, based on the results of early acceptability studies [12,15], and lower than several simulation studies found to achieve pandemic control [2,4,7]. However, the observed average adoption rate was close to 25%, which was a lower threshold for a suppressive effect of tracing apps on the infection curve in the study of Yasaka and colleagues [7], while Wymant and colleagues [11] highlighted that every 1% increase in app adoption leads to a meaningful decrease in infections. With regard to acceptability studies, the discrepancy might be explained by the gap between intention to use a tracing app and actual behavior, a phenomenon known as the intention–behavior gap [28]. Another explanation is the reliance on nonrepresentative samples and the use of digital surveys that might be preferred by people with positive attitudes toward digital apps [29].

App Design and App Adoption

Low adoption rates may also be the result of app design. To the best of our knowledge, this is the first study to investigate the association between app quality characteristics and actual adoption rates for national COVID-19 contact tracing apps. Bayesian regression analyses suggest that adoption rates can be increased by optimizing app quality characteristics. In our analysis, optimizing engagement features appeared to be associated with higher adoption rates, relative to strategies focused on functionality, aesthetics, or information quality. This finding has several implications. First, one promising strategy, from an app design perspective, appears to be to focus on engagement. Second, the app's level of engagement is a modifiable factor that can quickly be made the target of developers and public health actions. Other factors that are associated with the acceptance of tracing apps are either not or minimally modifiable, like potential users' age, gender, comorbidities, medical preconditions, level of education, household income, and trust in the government [12,30]. Third, it is not enough for tracing apps to trace and warn users properly. They should also be designed in a way that makes them interesting to use, stimulate repeated engagement with the app, offer customization and tailoring to personal preferences, allow user input, and provide feedback. Fourth, the predicted adoption rate in the case of optimal engagement was 40%. This potential increase is comparable or even larger than the 17% increase in adoption recently identified for monetary incentives for installing a COVID-19 contact tracing app [30]. Although it might be speculated that such different adoption-increasing interventions have an additional effect on adoption rates, a compensatory effect is possible. Finally, given the importance and potential impact on user rates of engagement-friendly app designs, this topic might deserve as much attention as data security and privacy issues.

App Adoption and Infection Rates

Another important finding of this study was the association between higher app adoption rates and lower infection rates. As association is not necessarily causation and confounders have to be considered, results should be discussed and interpreted from different perspectives. First, following the line of reasoning that using tracing apps leads to lower infection, it is of interest that the predicted adoption rate of 60% in this empirical observational study required achieving a zero infection rate that was between the adoption rates reported for simulation studies by Hinch and colleagues [4] and Xia and Lee [2]. Congruent with simulation studies, our results may indicate a beneficial effect of COVID-19 contact tracing apps, in terms of controlling the pandemic. Similarly, in a combined empirical observational and modeling study, Wymant and colleagues [11] found that app adoption as measured by downloads is an indicator of active use, and higher app use leads to lower infections. Interestingly, compared to the findings of Wymant and colleagues [11], in our study, we observed almost the same effect of app adoption, namely, a 1% increase in app adoption was associated with a 2.3% reduction in infections. Second, the association might also be explained by reverse causality. Growing infection rates (eg, in the beginning of the second wave) could trigger fear, which in turn may increase app adoption. Building on the well-investigated effects of fear on health behavior change, it is possible that some individuals adopted tracings apps as a consequence of an increased perceived threat. However, it was found that the effect of fear on behavior is rather limited [31], and sometimes fear could also lead to undesired consequences (eg, avoidance) [32]. Therefore, the association between adoption and infection may partly but not largely be explained by reverse causality. In line with our results on the importance of engagement-friendly app design, it is also possible that spikes in incidences may raise awareness for the existence of tracing apps and motivate individuals to download apps given the precondition that apps are perceived as beneficial and engaging. Third, it seems likely that third variables play a considerable role in explaining the overserved association, as we found no country using tracing apps as the only nonpharmaceutical measure. In fact, they are used in concert with a multitude of other nonpharmaceutical interventions [33]. For example, app users are also more likely to adhere to other public health measures, like adherence to social distancing recommendations, wearing a mask, and washing hands [30]. Accordingly, tracing app use probably represents one specific health behavior that is part of a broader set of protective attitudes and behaviors for self and others. Depending on the respective national policy, app use could facilitate and boost other measures (eg, testing as a consequence on exposure notification or results from an app symptom checker) [11]. In this case COVID-19 tracing apps could also be considered as a hub for a multitude of measures to fight the pandemic. Accordingly, it is difficult to disentangle the unique contribution of tracing apps, and interactions between measures should be considered when interpreting the considerable strong association between app adoption and infection.

Features and Recommendations

The features identified in this study using MARS as a guiding framework include those mentioned in prior reviews, such as self-monitoring, symptom checking [18,34], and contacts to helplines [35]. Although information or news related to COVID-19 represent broad categories [18,35], this study provides a more detailed description of existing features; for example, by differentiating between information content and purpose. Moreover, new features were observed like venue check-in, contact diary, and providing users control over tracing activity within the app. Generally, we observed a tendency toward multifeature COVID-19 contact tracing apps. Although apps evaluated earlier generally were limited to specific functionalities and features (eg, separate apps for tracing, symptom checking, and information dissemination [18]), apps reviewed later in 2020 seem to incorporate different features, thereby making the COVID-19 contact tracing app a broader hub for preventing COVID-19 transmission. Likewise, Collado-Borrell and colleagues [35] found that most apps had more than one purpose.

Based on best practice examples and the features observed in existing apps, recommendations were derived to foster an engagement-oriented app design (see Textbox 1).

Textbox 1. Recommendations for an engagement-oriented design for COVID-19 contact tracing apps.

- Strive for a multifeature tracing app and avoid a trace-and-warn functionality only app. Offer additional features that make daily living during the pandemic more convenient and secure (eg, by replacing paper contact lists with a digital venue check-in).
- Make the app interesting and relevant to stimulate daily use by providing useful information (eg, current infection rates; intensive care unit capacities; lockdown rules; and legal regulations pertaining to schools, work, commuting, etc)
- Make the app as regionally targeted as possible, indicating the region most relevant to each particular user's daily life, since public health measures (eg, regional lockdown rules) might only apply to certain regions. Matching the app's level of regionality to local applied public health measures might increase user acceptance and adherence.
- Address the need of relatedness and create a virtual community of people who are committed to protecting themselves and others (eg, by making sharing the app as easy and obvious as possible). Provide feedback on other users' protective behaviors toward self and others (eg, indicating the number of people using the app or sharing a positive test result).
- Include interactive and personalized features, like risk group and symptom checkers, and a symptom diary with individual feedback and recommendations. Such features encourage protective behaviors for both self and others, including early testing. They provide additional beneficial features to the user, stimulating frequent use, while personalization makes the app more relevant to users.
- Account for suboptimal connectivity between app users and extend the benefits of the digital solution to those not using a tracing app. For example, a venue check-in serves as a double safeguard, while a contact diary helps to inform those not using an app, thereby increasing the app's overall effectiveness.
- Reduce uncertainty and provide users with a clear picture of what will happen in case of infection. Succinctly and clearly describe what exactly the app will do if the user comes in contact with a person who is infected and how infected users can inform others.
- Adhere to current design trends to make the app visually appealing to as many as possible. Allow customized layouts or consider different graphic layouts for different target groups—like younger children, teenagers, and older adults—to better meet their needs and expectations. Prefer the currently widely employed tab navigation, hamburger menu, and card-based layout.
- Address the need of autonomy by providing users with full information and control over data. Settings should be customizable in both directions, allowing users to either pause and delete or share more detailed personal data than currently required.

Recommendations based on the reviewed best practice examples are consistent with other studies on features designed to increase user engagement with mobile health apps [36], especially those providing diary or note-taking functionality, health monitoring features, personalized feedback, personalized information matched to user characteristics (like their region of residence), providing autonomy through customizable settings, and screen design and navigation. Besides inspiration from existing apps and features, recommendations for further development could be driven by either the literature [36-38] or upcoming challenges over the course of the pandemic. First, rewards should be used more intensively (eg, for opening the app, sharing a positive test result with others, or recommending the app to others). Second, notifications and reminders on, for example, updated information or keeping a diary are likely to increase engagement with the app. Third, the perception of personal benefits is a facilitator of contact tracing and may be achieved by providing feedback on what the app has done for the user. Likewise, addressing collective responsibility is likely to facilitate engagement, for example, by providing feedback and rewards on what the user has done for the community and what the community has contributed to pandemic control, or by setting collective goals, for example, pertaining to the number of app recommendations or downloaded apps. Fourth, rewards and incentives could also be linked to the national health policy (eg, access to free of charge testing could be offered to users that received an alert or obtained critical results from a symptom checker). Likewise, using a venue check-in could be rewarded by easier access to social events. Such rewarding *health policy–related app characteristics* seem to have the potential to further increase app adoption. However, those health policy–related characteristics should be developed and implemented in close collaboration with lay publics such as civil society representatives, advocacy groups, and nongovernmental organizations [39]. Fifth, as greater interactivity and personalized features of tracing apps require more personal information (eg, information on regional lockdown rules require information on the user's location), it is even more important to implement the highest standards of transparency, data privacy, and control over personal data. Nevertheless, implementing those features seems important, as a recent study by Meier and colleagues [40] found that the perceived benefits of tracing apps are more important for the intention to use an app than privacy concerns from a user perspective.

As the percentage of people who have been vaccinated increases, two strategies are possible: tracing app promotion can be slowed or their functionality adapted to the changing situation. The latter should be considered an adaptive strategy, capitalizing on the observation that millions have downloaded the app and the premise that digital solutions to upcoming challenges could be useful. Such an adaptive strategy consists of slowly stepping down tracing features while stepping up the number of features designed to manage the pandemic's next phase, starting with the option of informing others if they have had contact with a vaccinated and, hence, potentially less infectious person [41] or one who has not yet been vaccinated. The adaptive strategy could include features designed to increase the rate of vaccinations, shifting the target behavior from tracing to undergoing vaccination. It might provide useful and personalized information; support users making appointments for health assessments, treatment, or vaccinations; report adverse or long-term effects; or aid users wanting to contact health care professionals. Designing adaptive tracing apps that are capable of changing their focus depending on a pandemic's life cycle appears feasible, of benefit to protecting public health, and worthy of further investigation.

Study Strengths

First, to evaluate quality characteristics, we systematically used an interdisciplinary approach by gathering a team of experts providing both digital health and human–computer interaction perspectives. All the apps were evaluated by four independent raters with two experts from digital health science and two from human–computer interaction research. Second, we included national apps from countries with similar identical legal regulations on voluntariness, privacy, and data protection. By the time this study was conducted, all countries adhered to the European GDPR or equivalent regulations. Therefore, confounding factors related to voluntariness, privacy, and data protection standards are unlikely to affect results to a greater extent. Nevertheless, differences in the integration of tracing apps into the respective health care system and national strategy to fight the pandemic exist.

Third, we investigated associations between app quality characteristics, actual national adoption rates, and infection rates. For the first time, results indicated a direct association between app quality characteristics and adoption rates that were, in turn, associated with lower infection rates. Although causality cannot be inferred from observational data, these results are consistent with prior simulation studies. Unless contradictory experimental data become available, our results should encourage app developers and policy makers to focus more on developing engagement-friendly designs for COVID-19 contact tracing apps to potentially enhance pandemic control. Third, prior reviews investigated all kinds of COVID-19–related apps, including apps with nontracing functionality (eg, only providing information to patients or health care professionals) and apps developed by both governmental and nongovernmental (eg, private technology) agencies [18,34,35]. This study focused entirely on COVID-19 contact tracing apps released by national health authorities, rendering the results especially important to public health officials and policy makers. Fourth, this study provides an estimation on potentially achievable adoption rates in case of optimal app design. For an optimistic outlook, the upper bound of the BCAR 80% highest density interval suggested an adoption rate of 61% for maximum MARS total scores and 54% for maximum engagement scores. However, these estimates also point to the necessity of additional measures beyond app-based digital contact tracing, as adoption rates are still below the thresholds indicated by simulation studies. Especially for younger children or older adults (eg, those with dementia), wearables may be an option to increase the overall adoption of digital contract tracing in the population [2].

Study Limitations

Several study limitations must be considered. First, app development is an ongoing and dynamic process. Consequently, this review merely provides a snapshot of the current status of app quality. The study is not a final report investigating the success or failure of specific national apps or digital contact tracing in general. Instead, it aims to support developers and public health policy makers striving to improve COVID-19 contact tracing apps, focusing on a predetermined set of important, yet limited, app quality characteristics evaluated by the MARS. Second, the number of downloaded apps was used as our measure of app adoption, and more precise measures exist, like the number of times a user opens the app, how long an app is used, and how many positive test results are shared. Moreover, we do not know the percentage of users who uninstalled the app. In one Swiss study, the uninstallation rate was 5% [29]. On the other hand, more precise measures conflict with data privacy, and the number of downloads was the only measure that was widely available to us. Third, item 15 of the MARS indicates that the correctness of information and scores might be biased toward more positive ratings. In cases of uncertainty, we assumed higher values, mainly because all the apps were released by credible sources: national health authorities. From a technical perspective, we were unable to determine if sensors worked properly and, therefore, if the information on contacts was reliable and accurate. From a medical perspective, COVID-19–related information was not reviewed by an independent board of virologists. However, we are not aware of any leading virologists who have criticized the information provided by any of the tracing apps we reviewed. Likewise, we had no information on diagnostic accuracy (ie, the sensitivity, specificity, and positive and negative predictive values of personal risk evaluations or symptom checkers). App developers should strive to include validated tools, since one study by Munsch and colleagues [42] found that some symptom checkers performed no better than random guessing, while some were highly sensitive, others highly specific, and only two both adequately sensitive and specific. Fourth, MARS, the instrument used in this study, does not include a section on privacy and security, a topic that was widely discussed in the media in the early phases of tracing app development [43]. Vokinger and colleagues [44] recently provided a framework for evaluating this issue in COVID-19 tracing apps.

Fifth, we focused exclusively on the characteristics of tracing apps and were unable to assess how these apps were embedded in the overall health care system or whether there were changes over time in the way tracing apps were embedded. Finally, we only assessed the international English version of each app and did not rule out the possibility that app versions in the language of the app's country of origin might include more information or features.

Conclusions

The member states of the European Union, associated countries, and former members with almost identical legal regulations on the voluntariness of use and data protection regulations serve as an excellent real-life laboratory for investigating tracing apps that have been released by the various national health authorities. All 21 national COVID-19 contact tracing apps that we evaluated demonstrated high levels of functionality, aesthetics, and information quality. Although, adoption rates were below the desired levels recommended by several simulation studies investigating the impact of digital contract tracing, while other research indicates that every additional percent app adoption is important. Our results suggest that engagement-friendly app design has the potential to gain those additional percentages. We found a lower level of an engagement-friendly app design, with substantial variations between the various apps. Moreover, app designs that raters considered more engagement-oriented were generally linked to higher app adoption rates, which makes developing more engagement-friendly app designs a promising target as countries strive to optimize their COVID-19 contact tracing apps. Several specific recommendations based on best practice examples were provided. The association we observed between higher app adoption rates and lower infection rates are consistent with predictions from simulation studies and—despite limitations—could indicate that COVID-19 contact tracing apps could contribute to flattening COVID-19 curves.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Overview on COVID-19 contact tracing apps considered for Mobile App Rating Scale rating.

[PDF File (Adobe PDF File), 136 KB - jmir_v23i6e27989_app1.pdf]

References

1. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
2. Xia Y, Lee G. How to return to normalcy: fast and comprehensive contact tracing of COVID-19 through proximity sensing using mobile devices. arXiv. 2020 Apr 27. URL: https://arxiv.org/abs/2004.12576 [accessed 2021-01-20]
3. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, C4591001 Clinical Trial Group. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. N Engl J Med 2020 Dec 31;383(27):2603-2615 [FREE Full text] [doi: 10.1056/NEJMoa2034577] [Medline: 33301246]
4. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective configurations of a digital contact tracing app: a report to NHSX. GitHub. 2020 Aug 10. URL: https://github.com/BDI-pathogens/covid-19_instant_tracing/blob/master/Report%20-%20Effective%20Configurations%20of%20a%20Digital%20Contact%20Tracing%20App.pdf [accessed 2021-01-20]
5. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. Lancet Public Health 2020 Aug;5(8):e452-e459 [FREE Full text] [doi: 10.1016/S2468-2667(20)30157-2] [Medline: 32682487]
6. Newzoo. 2020 smartphone users per region. URL: https://newzoo.com/key-numbers#mobile [accessed 2021-02-10]
7. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-peer contact tracing: development of a privacy-preserving smartphone app. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
8. Moreno López JA, Arregui García B, Bentkowski P, Bioglio L, Pinotti F, Boëlle PY, et al. Anatomy of digital contact tracing: role of age, transmission setting, adoption, and case detection. Sci Adv 2021 Apr;7(15):eabd8750. [doi: 10.1126/sciadv.abd8750] [Medline: 33712416]
9. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: 10.1016/S2589-7500(20)30184-9] [Medline: 32839755]
10. Kucharski AJ, Klepac P, Conlan AJK, Kissler SM, Tang ML, Fry H, CMMID COVID-19 working group. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. Lancet Infect Dis 2020 Oct;20(10):1151-1160 [FREE Full text] [doi: 10.1016/S1473-3099(20)30457-6] [Medline: 32559451]
11. Wymant C, Ferretti L, Tsallis D, Charalambides M, Abeler-Dörner L, Bonsall D, et al. The epidemiological impact of the NHS COVID-19 App. Nature 2021 May 12:1. [doi: 10.1038/s41586-021-03606-z] [Medline: 33979832]
12. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
13. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
14. O'Callaghan ME, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, et al. A national survey of attitudes to COVID-19 digital contact tracing in the Republic of Ireland. Ir J Med Sci 2020 Oct 16:1 [FREE Full text] [doi: 10.1007/s11845-020-02389-y] [Medline: 33063226]
15. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact tracing apps: predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]

16. Horstmann KT, Buecker S, Krasko J, Kritzler S, Terwiel S. Who does or does not use the 'Corona-Warn-App' and why? Eur J Public Health 2021 Feb 01;31(1):49-51 [FREE Full text] [doi: 10.1093/eurpub/ckaa239] [Medline: 33340328]
17. 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]
18. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for corona virus disease 2019 (COVID19) and the assessment of their quality using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
19. Messner E, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
20. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-12-23]
21. Emaasit D. Pymc-learn: practical probabilistic machine learning in Python. arXiv. 2018 Oct 31. URL: http://arxiv.org/pdf/1811.00542v1 [accessed 2021-01-20]
22. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
24. 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]
25. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
26. Sander LB, Schorndanner J, Terhorst Y, Spanhel K, Pryss R, Baumeister H, et al. 'Help for trauma from the app stores?' A systematic review and standardised rating of apps for post-traumatic stress disorder (PTSD). Eur J Psychotraumatol 2020;11(1):1701788 [FREE Full text] [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
27. Paganini S, Terhorst Y, Sander LB, Catic S, Balci S, Küchler AM, et al. Quality of physical activity apps: systematic search in app stores and content analysis. JMIR Preprints. Preprint posted online on January 5, 2021. [doi: 10.2196/22587]
28. Sheeran P. Intention—behavior relations: a conceptual and empirical review. Eur Rev Soc Psychol 2002 Jan;12(1):1-36. [doi: 10.1080/14792772143000003]
29. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of acceptance of COVID-19 proximity tracing apps in Switzerland: panel survey analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]
30. Munzert S, Selb P, Gohdes A, Stoetzer LF, Lowe W. Tracking and promoting the usage of a COVID-19 contact tracing app. Nat Hum Behav 2021 Feb;5(2):247-255. [doi: 10.1038/s41562-020-01044-x] [Medline: 33479505]
31. Kok G, Peters GY, Kessels LTE, Ten Hoor GA, Ruiter RAC. Ignoring theory and misinterpreting evidence: the false belief in fear appeals. Health Psychol Rev 2018 Jun;12(2):111-125. [doi: 10.1080/17437199.2017.1415767] [Medline: 29233060]
32. Stolow JA, Moses LM, Lederer AM, Carter R. How fear appeal approaches in COVID-19 health communication may be harming the global community. Health Educ Behav 2020 Aug;47(4):531-535. [doi: 10.1177/1090198120935073] [Medline: 32527161]
33. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. Nat Hum Behav 2020 Dec;4(12):1303-1312. [doi: 10.1038/s41562-020-01009-0] [Medline: 33199859]
34. Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, et al. COVID-19 mobile apps: a systematic review of the literature. J Med Internet Res 2020 Dec 09;22(12):e23170 [FREE Full text] [doi: 10.2196/23170] [Medline: 33197234]
35. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]
36. 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 09;22(12):e21687 [FREE Full text] [doi: 10.2196/21687] [Medline: 33295292]
37. Megnin-Viggars O, Carter P, Melendez-Torres GJ, Weston D, Rubin GJ. Facilitators and barriers to engagement with contact tracing during infectious disease outbreaks: a rapid review of the evidence. PLoS One 2020;15(10):e0241473 [FREE Full text] [doi: 10.1371/journal.pone.0241473] [Medline: 33120402]

38. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. Nat Hum Behav 2020 May;4(5):451-459. [doi: 10.1038/s41562-020-0887-9] [Medline: 32377018]
39. Blasimme A, Vayena E. What's next for COVID-19 apps? Governance and oversight. Science 2020 Nov 13;370(6518):760-762. [doi: 10.1126/science.abd9006] [Medline: 33184192]
40. Meier Y, Meinert J, Krämer N. Investigating factors that affect the adoption of Covid-19 contact-tracing apps. A privacy calculus perspective. PsyArXiv. Preprint posted online on February 10, 2021. [FREE Full text] [doi: 10.31234/osf.io/xk3t2]
41. Levine-Tiefenbrun M, Yelin I, Katz R, Herzel E, Golan Z, Schreiber L, et al. Decreased SARS-CoV-2 viral load following vaccination. medRxiv. Preprint posted online on February 8, 2021. [FREE Full text]
42. Munsch N, Martin A, Gruarin S, Nateqi J, Abdarahmane I, Weingartner-Ortner R, et al. Diagnostic accuracy of web-based COVID-19 symptom checkers: comparison study. J Med Internet Res 2020 Oct 06;22(10):e21299 [FREE Full text] [doi: 10.2196/21299] [Medline: 33001828]
43. Amann J, Sleigh J, Vayena E. Digital contact-tracing during the Covid-19 pandemic: an analysis of newspaper coverage in Germany, Austria, and Switzerland. PLoS One 2021;16(2):e0246524 [FREE Full text] [doi: 10.1371/journal.pone.0246524] [Medline: 33534839]
44. Vokinger KN, Nittas V, Witt CM, Fabrikant SI, von Wyl V. Digital health and the COVID-19 epidemic: an assessment framework for apps from an epidemiological and legal perspective. Swiss Med Wkly 2020 May 04;150:w20282. [doi: 10.4414/smw.2020.20282] [Medline: 32418194]

Abbreviations

BCAR: best-case adoption rate
GDPR: General Data Protection Regulation
ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
OLS: ordinary least squares
QR: Quick Response

Edited by G Eysenbach; submitted 16.02.21; peer-reviewed by V von Wyl, H Kondylakis; comments to author 09.03.21; revised version received 19.04.21; accepted 22.04.21; published 02.06.21.

Please cite as:
Kahnbach L, Lehr D, Brandenburger J, Mallwitz T, Jent S, Hannibal S, Funk B, Janneck M
Quality and Adoption of COVID-19 Tracing Apps and Recommendations for Development: Systematic Interdisciplinary Review of European Apps
J Med Internet Res 2021;23(6):e27989
URL: https://www.jmir.org/2021/6/e27989
doi:10.2196/27989
PMID:33890867

Original Paper

Quantifying Online News Media Coverage of the COVID-19 Pandemic: Text Mining Study and Resource

Konrad Krawczyk[1], DPhil; Tadeusz Chelkowski[2], MSc; Daniel J Laydon[3], PhD; Swapnil Mishra[3], PhD; Denise Xifara[4], DPhil; Benjamin Gibert[4], MSc; Seth Flaxman[5], DPhil; Thomas Mellan[3], PhD; Veit Schwämmle[6], PhD; Richard Röttger[1], PhD; Johannes T Hadsund[1], MSc; Samir Bhatt[3,7], DPhil

[1]Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark

[2]Department of Management in the Network Society, Kozminski University, Warsaw, Poland

[3]Department of Infectious Disease Epidemiology, MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, United Kingdom

[4]Nupinion, London, United Kingdom

[5]Department of Mathematics, Imperial College London, London, United Kingdom

[6]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark

[7]Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Konrad Krawczyk, DPhil
Department of Mathematics and Computer Science
University of Southern Denmark
Campusvej 55
Odense, DK-5230
Denmark
Phone: 45 6550 2387
Email: konradk@imada.sdu.dk

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/7/e31544/

Abstract

Background: Before the advent of an effective vaccine, nonpharmaceutical interventions, such as mask-wearing, social distancing, and lockdowns, have been the primary measures to combat the COVID-19 pandemic. Such measures are highly effective when there is high population-wide adherence, which requires information on current risks posed by the pandemic alongside a clear exposition of the rules and guidelines in place.

Objective: Here we analyzed online news media coverage of COVID-19. We quantified the total volume of COVID-19 articles, their sentiment polarization, and leading subtopics to act as a reference to inform future communication strategies.

Methods: We collected 26 million news articles from the front pages of 172 major online news sources in 11 countries (available online at SciRide). Using topic detection, we identified COVID-19–related content to quantify the proportion of total coverage the pandemic received in 2020. The sentiment analysis tool Vader was employed to stratify the emotional polarity of COVID-19 reporting. Further topic detection and sentiment analysis was performed on COVID-19 coverage to reveal the leading themes in pandemic reporting and their respective emotional polarizations.

Results: We found that COVID-19 coverage accounted for approximately 25.3% of all front-page online news articles between January and October 2020. Sentiment analysis of English-language sources revealed that overall COVID-19 coverage was not exclusively negatively polarized, suggesting wide heterogeneous reporting of the pandemic. Within this heterogenous coverage, 16% of COVID-19 news articles (or 4% of all English-language articles) can be classified as highly negatively polarized, citing issues such as death, fear, or crisis.

Conclusions: The goal of COVID-19 public health communication is to increase understanding of distancing rules and to maximize the impact of governmental policy. The extent to which the quantity and quality of information from different communication channels (eg, social media, government pages, and news) influence public understanding of public health measures remains to be established. Here we conclude that a quarter of all reporting in 2020 covered COVID-19, which is indicative of information overload. In this capacity, our data and analysis form a quantitative basis for informing health communication strategies along traditional news media channels to minimize the risks of COVID-19 while vaccination is rolled out.

(J Med Internet Res 2021;23(6):e28253) doi:10.2196/28253

KEYWORDS

text mining; COVID-19; infoveillance; sentiment analysis; public health

Introduction

The emergence of the novel coronavirus SARS-CoV-2 and the resultant disease COVID-19 has resulted in an estimated 2.4 million deaths [1,2]. Due to the initial lack of pharmaceutical measures targeting COVID-19, many governments resorted to nonpharmaceutical interventions (NPIs) to control the spread of the pandemic [3,4]. The introduction of NPIs, such as social distancing, mask-wearing, or so-called *lockdowns*, significantly reduced SARS-CoV-2 transmission [5-8]. Therefore, in the absence of effective treatments or widespread rollout of vaccines, NPIs remain an important tool in COVID-19 control [9].

The effectiveness of NPIs is dependent on population-wide adherence to government mandates (eg, social distancing rules, stay-at-home orders, and mask-wearing). Adherence, in turn, is dependent on society’s perception of such measures [9,10], which are shaped by print and digital media. Since adherence to NPIs is linked to public understanding of the guidelines, it is crucial for news sources shaping such knowledge to effectively expound the rules to maximize public response. As evidence accumulates, it is expected that guidelines will shift and be clarified. In the digital age, one of the primary information sources for society is online news [11,12]. Effective communication on the current state of the pandemic and prevention guidelines affects how society adheres to and responds to NPIs and, therefore, influences the severity of the pandemic.

News articles have been previously shown to be an effective way of tracking disease outbreaks, with services such as HealthMap contributing to detecting and tracking disease outbreaks [13,14]. Even if only by virtue of its unprecedented scale, news media has been particularly important in the COVID-19 pandemic [15]. There have been an estimated 38 million English-language articles on COVID-19 [16]. It has further been demonstrated that misinformation on COVID-19 has been widespread and influential, both on social media [17] and in traditional news sources [16]. This can lead to an overload of information that hinders societal response to the pandemic [18].

A particular difficulty of the pandemic is its emotional toll, both from the disease itself and from social distancing measures [19-21]. Emotional toll can be investigated quantitatively by sentiment analysis, which calculates emotional polarization of text from negative, through neutral, to positive. Three previous studies have attempted to quantify emotional toll using sentiment analysis of social media conversations on COVID-19 [22-24]. Counterintuitively, given pandemic subject matter, all three studies of sentiments from COVID-19 conversations on social media showed a higher proportion of positive rather than negative emotions. In contrast, analysis of COVID-19 headlines from 25 English news media sources indicated that 52% of them evoked negative emotions, 30% evoked positive emotions, and 18% evoked neutral emotions [25]. News media can shape behavior toward the pandemic and adherence to control measures. Therefore, extensive negative coverage or contradictory information (ie, information overload) could have detrimental effects to both the mental health of individuals and how effectively society responds to control measures [26].

An earlier study on COVID-19 news coverage by Evanega et al [16] sourced 38 million English-language COVID-19 articles by keyword search from LexisNexis, which indexes 7 million sources. Quantifying the extent of COVID-19 information requires normalizing the absolute number of articles with respect to the number of contributing sources. Likewise, sentiment of COVID-19 coverage also needs to be analyzed in the context of overall negativity of consumed information. Aslam et al [25] analyzed 141,208 COVID-19 headlines, showing that 52% carried negative sentiment. Similar results were reported by Chakraborty and Bose, who collected COVID-19 news articles from GDELT (Global Database of Events, Language, and Tone) and found that pandemic coverage was mostly associated with negative sentiment polarization [27]. Though informative, these studies did not contrast COVID-19 sentiment distribution with the sentiment distribution of the sources they originated from. By contextualizing the sentiment distribution of pandemic reporting within that of the overall coverage, it is possible to draw meaningful conclusions on whether the amount of COVID-19 information is indeed more negatively polarized than what news media consumers are exposed to.

To address the above issues, we collected over 26 million articles from the front pages of 172 major online news sources from 11 countries and compiled these into a reusable database available at SciRide [28]. Firstly, we investigated trends in COVID-19 news with respect to all articles that appeared on the front pages in 2020. Secondly, we analyzed whether articles on COVID-19 were more sentiment-polarized compared with other articles. Finally, we analyzed the leading subtopics in COVID-19 coverage and assessed their sentiment polarization. Overall, our work aimed to elucidate the volume and content of news coverage of COVID-19 in traditional media as a basis for data-driven discussion regarding policy communication in the pandemic.

Methods

Curation of a Database of Front-Page News Articles

To assess coverage of COVID-19 and the sentiment it evoked, we analyzed the landing pages from major online news sources in countries with robust media presence. We selected the major online news sources from 11 countries: the United States, the United Kingdom, Canada, Australia, New Zealand, Ireland, Germany, France, Italy, Spain, and Russia. We included an additional *international* category to better reflect the global focus of certain online news sources.

For each country, the major online news sources were identified by reference to profiles in BBC Media, which is an authority in curating global news source information and lists of news sites with the most traffic and is curated by SimilarWeb. Focusing on major national online news sources, as defined by online visibility, captures some of the main sources in shaping societal knowledge and opinions [29]. It should be noted that the focus on online news sources excludes the impact of social media, epistemic communities, and other influences on public perception. However, due to their depth of penetration, political heterogeneity, and reliability, major online news sources provide an excellent proxy of overall public perception.

For each online news source, we collected the archived front-page snapshots dating back to 2015 via Internet Archive [30], cutting off coverage in 2020 at October 15. We used the pre-2020 articles to fine-tune the accuracy of article collection and provide statistics on reporting of certain topics pre–COVID-19. Each front page was sourced for potential news items using a custom-based pipeline we developed (Section 1 in Multimedia Appendix 1). We defined each article as the combination of the metadata elements title and description—sometimes referred to as headline and subheading—that are broadly akin to titles and abstracts in scientific publications [31]. Such metadata are reasonably standardized among online news sources and they offer a headline-like summary of the article, typically designed for sharing on social media, making them suitable for topic detection and sentiment analysis. In total, we collected 26,077,939 articles from front pages of 172 online news sources (Table 1), with the full list of contributing sources in Table S1 in Multimedia Appendix 1.

Contemporary news sites rapidly change their content, which is spread through multiple sections, so it is difficult to gauge the level of attention received by particular articles. Front pages of news sites should be reliable reflections of the information many users are exposed to, since they are the main points of entry. This is opposed to other article collection strategies, such as RSS (Really Simple Syndication) or downloading the entire website content, that can provide limited control over assessing how many people have actually read any given article [16,32]. Focusing our efforts on articles from landing pages of major online news sources enabled us to assess the number of COVID-19–related articles that a large proportion of online news consumers were exposed to.

Table 1. Number of online news sources and collected articles per country.

Country	Online news sources (N=172), n (%)	Collected articles (N=26,077,939), n (%)
Canada	13 (7.5)	1,269,200 (4.8)
Australia	8 (4.6)	1,124,859 (4.3)
Italy	13 (7.5)	1,526,521 (5.8)
The United Kingdom	21 (12.2)	4,977,792 (19.0)
The United States	33 (19.1)	4,388,383 (16.8)
France	9 (5.2)	1,951,608 (7.4)
Germany	18 (10.4)	2,348,403 (9.0)
Ireland	8 (4.6)	905,598 (3.4)
International	6 (3.4)	462,989 (1.7)
New Zealand	5 (2.9)	651,050 (2.4)
Russia	19 (11.0)	3,348,825 (12.8)
Spain	19 (11.0)	3,122,711 (11.9)

Topic Models

For each article we extracted, we analyzed the text content of the metadata title and description to determine whether the article could be associated with one of the following topics: cat, sport, Merkel, Putin, Johnson, Biden, Trump, cancer, climate, or COVID-19. Non–COVID-19 topics were selected to provide a reference with large expected volumes of coverage (ie, politicians) that cover a wide range of sentiments (e g, *cat* as nonnegative and *cancer* as negative). Each topic was identified on the basis of the keywords presented in Table 2. The only normalization we applied to the words for topic identification was case folding, otherwise the words were not stemmed nor lemmatized. Topics that were used solely for sentiment analysis—cat, sport, climate, and cancer—were not identified for non-English online news sources.

We chose keywords for each topic in order to maximize the precision of topic identification. Because we focused on titles and descriptions, mentions of specific keywords here made it unlikely that they were only tangentially relevant to the article at hand (eg, explicit mentions of politicians' names). In the case of COVID-19, we tested the extent to which our topic detection misclassified the topic by identifying COVID-19 articles in the pre–COVID-19 era of 2015 to 2019. Out of 21,693,591 articles, only 7375 (0.03%) were erroneously identified as related to COVID-19, demonstrating high precision of the selected keywords. In the majority of cases, misclassifications stemmed from mentions of lockdowns, which were chiefly gun related, but in the case of the United Kingdom, they were even related to a seagull attack on a school. Subsequent subtopic identification stratified the different threads of COVID-19 coverage, providing a wider set of keywords.

Table 2. Keywords employed for topic detection.

Topic	Keywords by language					
	English	German	French	Spanish	Italian	Russian
COVID-19						
	coronavirus	coronavirus	coronavirus	coronavirus	coronavirus	коронавирус
	covid	covid	covid	covid	covid	covid ковид
	lockdown	lockdown	lockdown couvre-feu	lockdown confinamiento	lockdown contenimento	lockdown локдаун
	quarantine	quarantäne	quarantaine	cuarantena	quarantena	карантин
	pandemic	pandemie	pandémie	pandemia	pandemia	пандемиа
	N/A[a]	corona-	N/A	N/A	N/A	N/A
Merkel	merkel	merkel	merkel	merkel	merkel	merkel меркел
Trump	trump	trump	trump	trump	trump	trump трамп
Biden	biden	biden	biden	biden	biden	biden байден
Johnson	boris johnson	boris johnson	boris johnson	boris johnson	boris johnson	boris johnson борис джонсон
Putin	putin	putin	putin	putin	putin	putin путин
Climate						
	global warming	—[b]	—	—	—	—
	climate change	—	—	—	—	—
	climate crisis	—	—	—	—	—
Cat						
	cat	—	—	—	—	—
	kitten	—	—	—	—	—
Sport						
	baseball	—	—	—	—	—
	major league	—	—	—	—	—
	champion's league	—	—	—	—	—
	football	—	—	—	—	—
	nfl	—	—	—	—	—
	premier league	—	—	—	—	—
	basketball	—	—	—	—	—
	soccer	—	—	—	—	—
	nba	—	—	—	—	—
Cancer	cancer	—	—	—	—	—

[a] N/A: not applicable; this keyword, which is specific to the German language because of its compound nature, was only found in German news sources.

[b] The topics *climate*, *cat*, *sport*, and *cancer* were not identified in non–English-language online news sources, as these were solely employed for sentiment analysis.

We further identified subtopics within COVID-19 news coverage by a similar keyword-based approach. Since many subtopic words can have several forms (eg, dead, died, and dies), we stemmed the words associated with each subtopic (eg, healthy and healthier are both stemmed to health); these are presented in Table 3. An article was identified as pertaining to a subtopic if, after stemming its title and description, a token corresponding to a stemmed keyword in Table 3 was identified.

Table 3. COVID-19 news subtopics.

Subtopic[a]	Stemmed keywords
Case	case
Crisis	crisi
Death	die, death
Disease	diseas
Distancing	distanc
Fear	fear
Health	health
Home	home
Hospital	hospit
Infection	infect
Isolation	isol
Lockdown	lockdown
Mask	mask
Outbreak	outbreak
Quarantine	quarantin
Spread	spread
Symptom	symptom
Test	test
Treatment	treatment
Vaccine	vaccin

[a] Each of the subtopics was identified by the stemmed keywords (ie, stemming).

Sentiment Analysis of News Articles Using Vader

We used a well-established sentiment analysis tool, Vader [33], to identify emotionally polarized content, which has previously been applied to news media. It is suitable for short snippets of text, such as the titles and descriptions in our metadata. For a given piece of text, Vader provides a compound score between –1 and 1, with –1 being entirely negative, 0 being neutral, and 1 being entirely positive. For instance, the phrase “I find your lack of faith disturbing” offers a Vader score of –0.42, whereas the phrase “I find your lack of faith encouraging” gives a score of 0.5994. In our case, a sentiment score for a single article consists of a Vader compound score for the concatenation of the article title and description.

Novel topics are associated with many subject-specific keywords and phrases (eg “social distancing” and “lockdown” for COVID-19). Applying sentiment analysis to text with novel keywords can result in software being unable to correctly annotate polarization. We assessed Vader sentiment annotations on articles identified as one of the subtopics in Table 3. This revealed an artifact of the tool, wherein “positive coronavirus test” was labeled as emotionally polarized in the positive direction by virtue of the word “positive.” In order to mitigate the effect of this subject-specific misclassification, the words “positive” and “negative,” for symmetry, were removed from articles related to coronavirus testing prior to applying Vader annotation.

Estimating Topic Polarization: Relative Sentiment Skew

We examined whether coverage of a given topic was more emotionally polarized than another topic by contrasting their sentiment distributions. Directly comparing topic sentiment distributions between different online news sources is not sound. Different online news sources can be more sensational and negative or toned-down and neutral, which gives radically different sentiment distributions. To address this issue, we calculated whether specific topic coverage was more negative, positive, or neutral relative to other articles within a particular online news source.

For each article “*a*” (ie, that article’s title and description metadata) in English-language online news sources, we calculated the Vader compound sentiment score *sent(a)*. For all 2020 articles and topics, in a given online news source, we calculated the mean of the Vader compound sentiment scores, denoted as $\mu_{ONS,TOPIC}$ (online news source [ONS]; equation 1). As a reference statistic for the distribution of sentiment scores not relating to a topic, we calculated the mean Vader score of the articles from a given online news source that were not identified as a given topic, denoted as $\mu_{ONS,TOPIC}$ (equation 2):

where *ONS* is a particular online news source, *TOPIC* is a topic from Table 2 or Table 3, *ONS(TOPIC)* is the set of articles on a given topic from a particular online news source, and $|ONS_{TOPIC}|$ is the total number of articles on a given topic in that online news source. A set of articles not identified as a given topic from a particular online news source is denoted as ONS_{TOPIC}.

For each topic in each online news source, we calculated the relative sentiment skew ($rsskew_{ONS,TOPIC}$) between topic mean sentiment and the mean of all other articles in the given online news source (equation 3).

Relative sentiment skew is designed to indicate whether the sentiment score distribution of a particular topic is negatively or positively polarized, compared to other articles, for example, where a topic has one positive (score +1) and two negative articles (each with score −1), and there are seven other nontopic articles that are all positive (each with score +1). In this instance, the relative sentiment skew metric $rsskew_{ONS,TOPIC}$ is –1/3 – 7/7 = –1.33, which indicates greater negativity. Note, we do not account for sample size variation, as the denominator is generally very large (ie, in the thousands).

Results

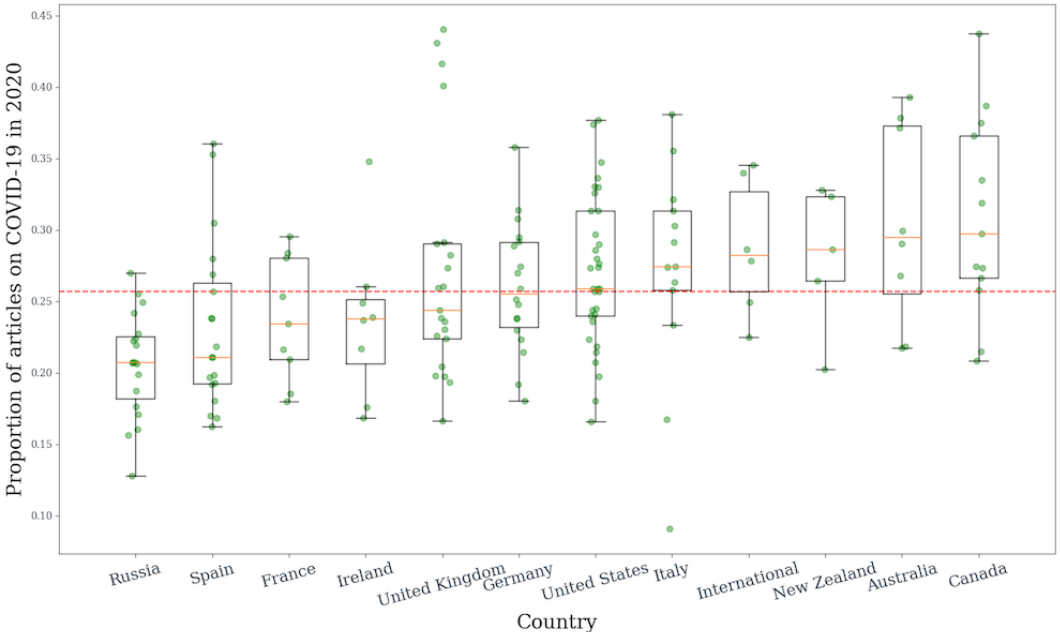
One-Quarter of 2020 News Coverage Was Pandemic Related, Suggesting Information Overload

We estimated the extent of COVID-19 coverage in online news media by identifying articles relating to the pandemic and comparing this number to the total number of articles between January and October 2020.

For each online news source, we performed topic detection, categorizing each article title and description as relating to the coronavirus—topic called COVID-19—if the title and description contained any keyword of a specific set, given in Table 2. Keywords for this simplified topic detection model were chosen to maximize the precision and accuracy of content identification in order to avoid cross-contamination with other topics. In English, these keywords included *covid*, *coronavirus*, *lockdown*, *quarantine*, and *pandemic*, but not, for example, *hospital* and *death*. The keywords were adjusted for the six languages that we used in this study: English, German, French, Spanish, Italian, and Russian (Table 2).

COVID-19 featured in 25.3% of the news articles (1,135,561/4,477,867) across the online news sources (Figure 1). While this proportion varied between countries, it was consistently large, with the lower bound at 20% and the upper bound at 30%. Thus, even using our relatively simple topic detection model, we were able to demonstrate that 2020 online news coverage was dominated by COVID-19.

Figure 1. The extent of coronavirus coverage in 2020. We calculated the proportion of all COVID-19 articles as the proportion of all front-page articles. Proportions were calculated for each online news source separately and then aggregated at the national level. The green points represent the individual coverage of each online news source. The yellow line in each box represents the median; the upper and lower whiskers represent the 75th and 25th percentiles, respectively. The red dotted line indicates mean proportion across all online news sources.



To provide a point of reference to coverage of COVID-19, we identified other global topics with regular and topical media presence (Table 2). We selected topics such as Donald Trump (Trump), Joe Biden (Biden), Boris Johnson (Johnson), Angela Merkel (Merkel), and Vladimir Putin (Putin). Unsurprisingly, mentions of politicians are most common in their home countries

(Figures S1-S5 in Section 2 of Multimedia Appendix 1) (eg, Vladimir Putin in Russia). Nonetheless, each politician received an order of magnitude less media attention than COVID-19 in 2020. Specifically, mean coverage of COVID-19 across 11 countries was 25.66%, whereas politicians received 2.50% for Trump, 0.45% for Biden, 0.18% for Putin, 0.17% for Johnson, and 0.09% for Merkel. Higher coverage of COVID-19 was also reflected at the national level. Although 2020 was an election year in the United States, *Trump* received a mean of 15.29% of 2020 media mentions in US online news sources, as opposed to 25.91% of mean mentions for COVID-19. Furthermore, US articles in 2020 mentioning both Trump and COVID-19 accounted for a mean coverage of 3.82%.

To provide temporal perspective on the media attention of COVID-19, we plotted the proportions of global coverage in the period from January to October 2020 (Figure S6 in Section 2 of Multimedia Appendix 1). COVID-19 attention in 2020 spiked between March and May, coinciding with many countries following the Chinese strategy of lockdown and other distancing measures. Proportion coverage leading up to the second COVID-19 wave in Europe and the United States did not reach levels seen from March to May but did stay above 20% in the regions we considered. These results provide a quantification of the extent of media attention received by COVID-19 across countries and languages, reflecting the global and protracted impact of the pandemic.

COVID-19 News Sentiment Analysis Suggests Heterogeneity of Coverage

The emotion that news coverage of COVID-19 evokes is an important factor in a society’s response to the pandemic [25]. We addressed this issue by sentiment analysis contrasting emotional polarization of COVID-19 news with that of certain reference topics and all non–COVID-19 articles for each online news source. Such contrast allows us to determine whether sentiment distribution of COVID-19 coverage was polarized compared to reference topics and other online news for a given online news source.

To quantify the emotional content of news article text, we employed Vader, which has been previously applied to news article analysis [33,34] (see Methods section). For each online news source whose primary language was English (91/172, 52.9%), we created Vader annotations for both the title and description of each of its articles. We grouped articles by their annotated topics (Table 2) to provide a comparison with COVID-19 sentiments. Politicians—Merkel, Trump, Johnson, Biden, and Putin—were used as reference points for subjects with frequent coverage. We selected four additional topics to offer intuitive reference points on the positive and negative sentiment spectrum: cat, sport, climate, and cancer. The topics *cat* and *sport* were used, as they would not necessarily be associated with negative sentiments. Likewise, the topics of *climate*—identified by the key phrases global warming, climate crisis, and climate change—and *cancer* were used as references that we would expect to be associated with negative emotions. Altogether, individual sentiment annotations for each online news source were grouped by one of the topics: cat, sport, Merkel, Johnson, Biden, Trump, COVID-19, and cancer.

For each online news source and topic, we calculated the relative sentiment skew statistic ($rsskew_{ONS,TOPIC}$; see Methods section), which measures the polarization of a given topic within an online news source. We noted how many topics had a negative or positive relative sentiment skew value (Table 4). For the nonpolitician non–COVID-19 topics (ie, cat, sport, and cancer), skew was in the expected direction, suggesting that they are appropriate references for assessing sentiment of COVID-19 articles. We noted negative relative sentiment skew values for 74 out of 91 (81%) English-language online news sources. Nonetheless, this observation cannot be taken as evidence of significant negative polarization, as these relative sentiment skew values are not substantially different than $rsskew=0$, which indicates no polarization. The mean relative sentiment skew for COVID was –0.04 (SD 0.07) (Table 4). Since cancer and COVID-19 are both diseases and exert a large burden on public health, one might expect their sentiment distributions from online news to be similar. However, COVID-19 sentiment distribution was not as extreme as that of cancer, which was 100% negative per online news source, and had a mean relative sentiment skew of –0.53 (SD 0.12). Perhaps surprisingly, the sentiment distribution for COVID-19 articles was more akin to coverage of climate, which was *a priori* expected to be negative, akin to cancer, or to subjects covering heterogenous topics by virtue of their wide-ranging implications for society, such as politicians (Figure 2). This, however, can be an indicator of topics being intertwined: since heads of state are responsible for the pandemic response, they can be expected to be mentioned in relation to COVID-19.

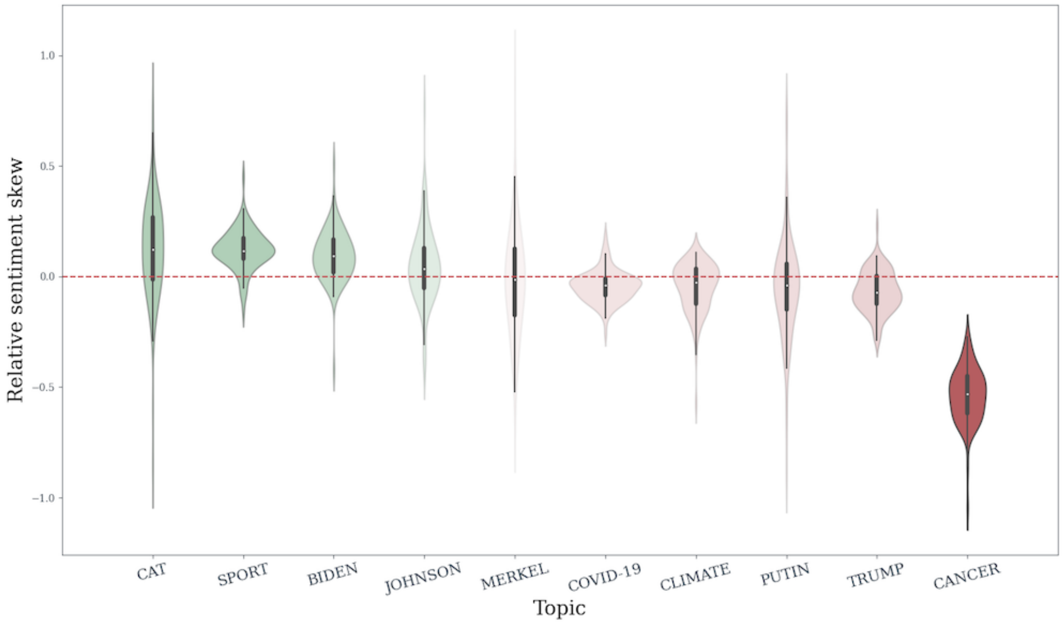
Table 4. English-language online news sources[a], with positive (≥0) or negative (<0) relative sentiment skew of 2020 articles on a given topic.

Topic	Positive online news sources, n (%)	Negative online news sources, n (%)	Relative sentiment skew, mean (SD)	Total articles[b], n
Cat (n=87)	64 (74)	23 (26)	0.12 (0.23)	2746
Sport (N=91)	84 (92)	7 (8)	0.12 (0.08)	63,155
Biden (n=90)	75 (83)	15 (17)	0.09 (0.11)	38,949
Johnson (n=90)	57 (63)	33 (37)	0.04 (0.17)	22,613
Merkel (n=79)	38 (48)	41 (52)	–0.01 (0.25)	2011
COVID-19 (N=91)	17 (19)	74 (81)	–0.04 (0.07)	589,701
Climate (N=91)	38 (42)	53 (58)	–0.04 (0.11)	7195
Putin (n=88)	33 (38)	55 (63)	–0.05 (0.23)	5179
Trump (N=91)	24 (26)	67 (74)	–0.06 (0.09)	157,702
Cancer (N=91)	0 (0)	91 (100)	–0.53 (0.12)	9548

[a] We had 91 English-language online news sources in total; however, in cases where it was impossible to identify a certain topic in a given source, it was left out.

[b] The total number of articles we identified as being associated with a given topic across all online news sources.

Figure 2. Relative sentiment skew (*rrskew*) of COVID-19 coverage. Each article title and description from each English-language online news source (ONS) received a Vader sentiment compound score between –1 and 1 (most negative and most positive, respectively). We noted the difference in mean sentiment for a specific topic and mean sentiment for other 2020 articles in a given online news source ($rsskew_{ONS,TOPIC}$; see Methods section). The density of the relative sentiment skew is plotted for each topic. Distributions are colored green if their relative sentiment skew was predominantly positive or red if predominantly negative (Table 4). Intensity of the color is scaled by the distance from the red dotted line at 0, which indicates a lack of difference between topic sentiment and all other articles in a given online news source.



These results suggest that the sentiment of COVID-19 coverage in online news media is heterogeneous and is certainly not as clearly polarized as cancer, though the volume of coverage might play a role (Figure S7 in Section 3 in Multimedia Appendix 1). One explanation could be the all-permeating nature of the pandemic, where it becomes background to most reporting. Therefore, COVID-19 articles cannot all be categorized as fully negative, contrary to the expectation of pandemic subject matter. In fact, on average, they appear to not be polarized in either the positive or negative direction, especially when compared to reference topics. This suggests that coverage of COVID-19 was highly heterogeneous, with many themes contributing to the totality of messaging.

Highly Sentiment-Negative Subtopics Account for 16% of COVID-19 Coverage, Suggesting Emotional Pressure

We studied the text content of COVID-19–related title and description metadata to reveal the leading themes associated with heterogeneous pandemic reporting.

We investigated article subtopics and calculated the most commonly used words and bigrams (ie, consecutive combinations of two words) to demonstrate the most frequent mentions in COVID-19 coverage. For each of the 91 English-language online news sources, we calculated the ranks of single words and bigrams in articles that pertained to COVID-19. The articles were further subdivided within each online news source as negative (Vader score <–0.2; 247,542 articles), positive (Vader score >0.2; 192,643 articles), or all (any Vader score; 589,709 articles). This subdivision aimed to reveal whether certain keywords or bigrams were more frequently associated with differently polarized text. For each subdivision, we averaged the individual online news source ranks of each word and bigram found across all 91 online news sources. In Table 5 we present the top 20 highest-ranking words and bigrams across all 91 online news sources. Words and bigrams in Table 5 reveal many themes that are intuitively associated with the coronavirus, such as testing, vaccine, death, etc. In particular, negative articles had unique top words and bigrams that are intuitively associated with negative emotions. In the singletons, these were *death*, *crisi*, and *fear*, while in bigrams these were *covid_crisi*, *covid_death*, *coronavirus_death*, and *death_toll*—note that words are stemmed.

To calculate the news coverage proportion related to these top themes, we created a constrained set of COVID-19 subtopics, based on the highest-ranking words and bigrams from Table 5. We removed terms that pointed to nonspecific coronavirus coverage, such as *COVID*, *coronavirus*, *pandemic*, or *news*. We extended the list of subtopics to include those we did not find in Table 5 but considered as strongly related to COVID-19 coverage, such as *hospital*, *quarantine*, *symptom*, or *isolation*, with a full list of subtopics in Table 3. For each subtopic, we calculated the proportion of COVID-19 coverage per online news source (Figure S8 in Section 4 of Multimedia Appendix 1) and relative sentiment skew per online news source for the subtopic (Figure S9 in Section 4 of Multimedia Appendix 1). The means of coverage and sentiments per online news source are plotted against each other in Figure 3. The subtopics in Table 3 account for a mean of 67.14% of all COVID-19 articles across English-language online news sources. Of these, the top three are *case*, *lockdown*, and *death*, which account for a mean of 9.29%, 8.56%, and 8.08% of COVID-19 articles, respectively. Figure 3 and Figure S9 in Multimedia Appendix 1 suggest that out of *case*, *death*, and *lockdown*, only *death* carried a firmly polarized sentiment, with *case* and *lockdown* not being significantly skewed in either positive or negative directions. The words *fear*, *crisis*, and *death* unsurprisingly indicated substantial negative polarization (Figure 3).

We analyzed the impact on news sentiment in 2020 of the three most negative subtopics, *fear*, *crisis*, and *death* (Figure 3). For each online news source, we calculated the mean sentiment of all the 2020 articles after removing articles that mentioned one of those three topics. For all 91 online news sources, removing articles mentioning one of the top three negative topics resulted in a statistically significant—at the level of .05 with Bonferroni correction—shift toward mean positive sentiment (Section 5 in Multimedia Appendix 1). By contrast, removal of all the sentiment-heterogeneous (Figure 2) articles from all 2020 articles resulted in a significant shift toward mean positive sentiment in 40 out of 91 (44%) online news sources, a significant shift toward mean negative sentiment for 11 out of 91 (12%) online news sources, and no statistically significant result for 39 out of 91 (43%) online news sources. Altogether, articles mentioning *fear, crisis*, and *death* accounted for a mean of 16% of COVID-19 articles across 91 online news sources; due to their highly polarizing nature, they may play a significant role in shaping societal perception of the pandemic.

Of the three most negative topics, *fear*, *crisis*, or *death*, the latter was the most frequently mentioned for COVID-19 (Figure 3). In total, *death* was mentioned in the context of COVID-19 in 2.33% of all coverage in 91 English-language online news sources. All *death* mentions in 2020 in 91 English-language online news sources accounted for a mean coverage of 5.74%, whereas in the pre–COVID-19 period of 2015 to 2019, they accounted for 4.07%. Therefore, we can identify that *death* in the context of COVID-19 constituted a significant proportion of negatively associated coverage that appears to have contributed to overall death reporting in the news.

These results demonstrate that even though the overall coverage of COVID-19 in 2020 was not significantly polarized by sentiment, there was a nontrivial proportion of negative news that contributed to overall reporting negativity in 2020.

Table 5. Top words and bigrams in English-language countries.

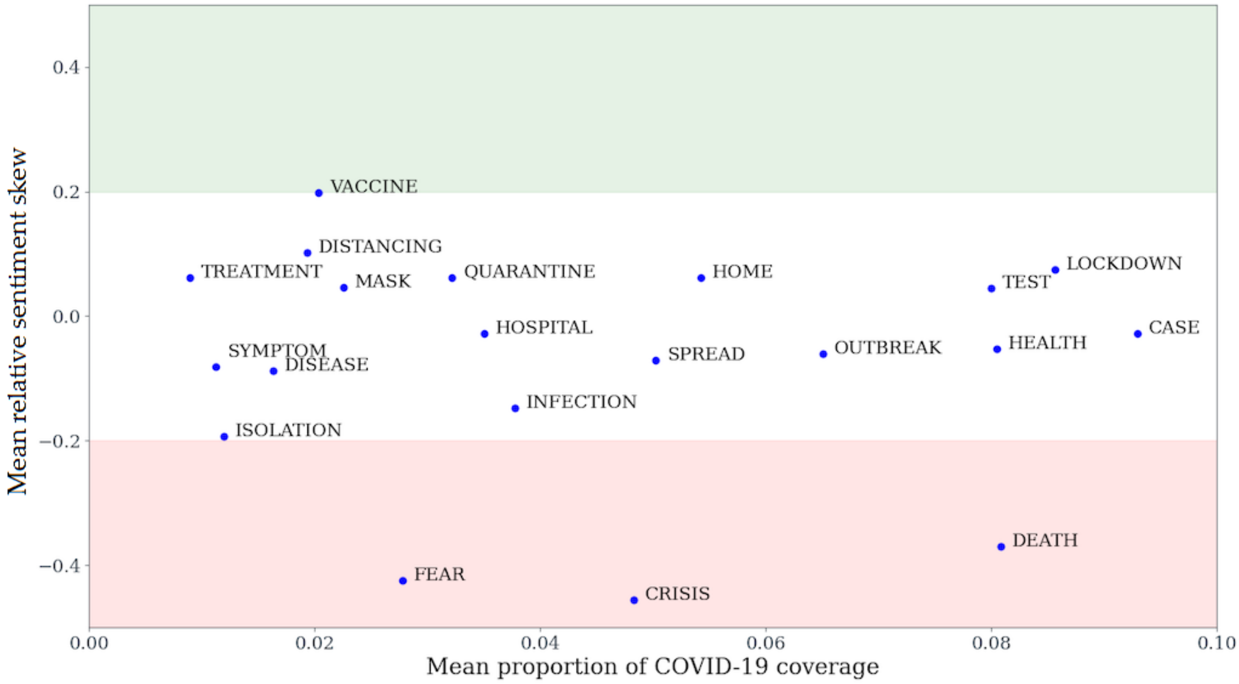
Rank	Words and bigrams by polarization[a]		
	Negative	All	Positive
Single words			
1	coronavirus	coronavirus	coronavirus
2	covid	covid	covid
3	pandem	pandem	pandem
4	new	new	new
5	peopl	peopl	help[b]
6	say	case	say
7	crisi[b]	say	peopl
8	health	health	test
9	case	test	health
10	death[b]	outbreak	case
11	outbreak	week	us
12	virus	us	home
13	test	virus	week
14	could	could	one
15	govern	day	time
16	us	one	day
17	countri	govern	govern
18	one	home	could
19	week	countri	work[b]
20	fear[b]	time	outbreak
Bigrams			
1	coronavirus_pandem	coronavirus_pandem	case_covid
2	coronavirus_crisi	case_coronavirus	coronavirus_pandem
3	coronavirus_outbreak	case_covid	posit_test
4	health_public	coronavirus_spread	health_public
5	posit_test	distanc_social	coronavirus_outbreak
6	case_coronavirus	health_public	case_coronavirus
7	coronavirus_spread	covid_test	covid_pandem
8	coronavirus_new	coronavirus_outbreak	distanc_social
9	case_covid	covid_pandem	coronavirus_lockdown[b]
10	distanc_social	coronavirus_new	coronavirus_spread
11	covid_crisi[b]	coronavirus_crisi	covid_test
12	covid_test	case_new[b]	covid_vaccin[b]
13	covid_pandem	covid_outbreak	home_stay
14	coronavirus_due[b]	posit_test	coronavirus_test[b]
15	covid_death[b]	minist_prime	covid_posit[b]
16	second_wave[b]	home_stay	minist_prime
17	death_toll[b]	first_time[b]	like_look[b]

Rank	Words and bigrams by polarization[a]		
	Negative	All	Positive
18	coronavirus_death[b]	around_world[b]	covid_outbreak
19	amid_coronavirus[b]	covid_spread[b]	coronavirus_new
20	two_week	two_week	coronavirus_vaccin[b]

[a] For each of the 91 English-language online news sources, we calculated the most common words and bigrams and grouped these by Vader scores: >0.2 for positive, <–0.2 for negative, and any score for all. We averaged the ranks of words and bigrams across all online news sources, and here we present the top 20 for each subdivision. The words in the table are stemmed.

[b] These entries indicate elements that can be found in the top 20, only in the specific subdivisions of *positive*, *all*, or *negative*.

Figure 3. COVID-19 subtopic coverage and sentiment means. We calculated the mean coverage and mean sentiment of each subtopic. Coverage is expressed as the mean of ratios of subtopics in a given online news source against all COVID-19 articles in the same online news source. Sentiment is the mean of the subtopic relative sentiment skew for all online news sources. The shaded areas illustrate regions with relative sentiment skew above 0.2 (green), between 0.2 and –0.2 (white), and below –0.2 (green).



Discussion

In this work, we compiled the largest data set on COVID-19 news, collating over 26 million articles from the front pages of 172 major online news sources from 11 countries. We have made this database publicly available at SciRide [28]. We first investigated trends in COVID-19–related news with respect to all front-page articles in 2020. We next used sentiment analysis to determine whether COVID-19 coverage was more polarized than other topics. Finally, we analyzed the leading subtopics in COVID-19 coverage and assessed their sentiment polarization. We demonstrated that 25% of front-page articles in traditional news media between January and October 2020 concerned COVID-19. Sentiment analysis demonstrated that pandemic coverage cannot be simply categorized as negatively polarized by virtue of disease association, pointing to heterogenous reporting. However, there was an increased incidence in reporting of negatively associated topics, in particular concerning *death*. Our results provide a data-driven foundation on policy communication surrounding distancing measures.

NPIs are drastic measures that reduce casualties before vaccinations and/or treatment regimens become widespread. Such methods, however, are only effective with societal adherence. Information received by a population in times of a pandemic shapes collective adherence to policies introduced to stem its spread. Currently, the internet is the primary source of health information for people in developed countries [35].

Comprehensive analysis of COVID-19 information received by a population would require thorough analysis of all possible internet news sources and all users' exposure to each piece of information received. The online ecosystem is extremely heterogeneous, with channels of information discovery spread across traditional news sites, blogs, social media, and many others. Within each of such platforms, information itself can take different forms (eg, text length and format). How users interact with the information also has a great effect on the amount of attention a given piece of information receives and the degree of influence it has (eg, extent of sharing on social media or a more visible position on a website). Analysis of COVID-19 information from all online sources is not tractable.

Direct access to major news sites accounts for 76% of media consumption online [29]. The landing pages of such outlets implicitly capture articles that might have been seen by online users. Thus, our analysis of content from front pages of major news sites should encompass a significant proportion of sources shaping knowledge of the pandemic covering reporting across different languages and geographies. In total, we curated a data set of 26 million articles from 172 major web traffic–generating online news sources in 11 countries.

We identified COVID-19–relevant articles as well as a selection of other topics to serve as reference points for both coverage volume and sentiment analysis. As a standardized common denominator among articles, we analyzed the metadata titles and descriptions, where the main subject matter can be expected to be referenced. We employed a facile topic identification method using a limited set of keyword mentions. We chose a limited number of keywords to make it unlikely that an article would not make the corresponding topic its subject matter if it were referenced in its metadata title and description (eg, politician’s name). This avoided the caveat of tangential references to certain topics mentioned in the full article body or ambiguities that might arise by using more sophisticated topic modeling algorithms [36]. Unlike more complex topic modeling methods, or even using a wider set of keywords, our approach did not capture much more subtle references to these topics, and so we will have underestimated total coverage.

Nonetheless, even using our simple approach, we still identified a nontrivial number of COVID-19 articles on the front pages of our online news sources. We estimate that a mean of 25% of our sample of front-page articles from 11 countries in 2020 mentioned COVID-19 in their titles and descriptions. Our method had reduced topic identification recall by not accounting for more subtle references to COVID-19, and the totality of the articles was certainly contaminated by retrieval of erroneous links that were not actual news articles. Therefore, the actual proportion of articles on the front pages of news sources referencing COVID-19 might have indeed been higher. We envision that the amount of reporting on a topic of general interest like COVID-19 needs to be balanced. Too little information might leave the population underinformed and ill-equipped to respond appropriately. Too much coverage risks obscuring information that is crucial for individuals to understand the pandemic and how to stay safe.

Reporting on the pandemic could have wider implications than only its basic informative function. It is unknown what effect regular reporting on cases, casualties, and containment methods could have on adherence to distancing rules or mental health [19,37-40]. Though current sentiment analysis methods fall short of identifying complex nuance, they offer a good approximation for the position of text on the emotional spectrum (ie, negative, neutral, or positive). By employing sentiment analysis, we found that overall COVID-19 reporting was not markedly polarized in either a positive or negative direction, as opposed to cancer. It is contrary to what might be expected by virtue of the pandemic subject matter, suggesting heterogeneity in reporting. Such heterogeneity might be due to the sheer scale of the pandemic, the consequences of which have permeated much of everyday life. Nevertheless, we found that negatively polarized COVID-19 articles mentioning death, fear, and crisis accounted for 16% of pandemic articles, with death being most widely referenced. Such a nontrivial volume of negatively associated articles significantly skews the sentiment of 2020 reporting toward the negative direction.

Our results offer a quantification of COVID-19 reporting that substantiates widespread qualitative observations (eg, the pandemic received an unprecedented amount of media attention). Our analysis offers insights for shaping discussion on health communication in order to maximize the effect of control policies. Our retrospective analysis of health communication during the first two waves indicates signs of information and emotional overload that might have obscured understanding of policy. We hope that our findings will inform how best to communicate so as to minimize the risk of subsequent waves while vaccination regimens are introduced.

Acknowledgments

SB would like to acknowledge the Academy of Medical Sciences, the National Institute for Health Research (NIHR) Biomedical Research Centres, and UK Research and Innovation. DL acknowledges funding from the Vaccine Efficacy Evaluation for Priority Emerging Diseases project (grant NIHR: PR-OD-1017-20002) from the NIHR. We would also like to acknowledge Microsoft AI for Health and Amazon AWS for computing resources.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 5402 KB - jmir_v23i6e28253_app1.docx]

References

1. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020 May 01;368(6490):489-493 [FREE Full text] [doi: 10.1126/science.abb3221] [Medline: 32179701]
2. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. World Health Organization. 2021. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2021-02-25]
3. Hyafil A, Moriña D. Analysis of the impact of lockdown on the reproduction number of the SARS-Cov-2 in Spain. Gac Sanit 2020 May 23:1-6 [FREE Full text] [doi: 10.1016/j.gaceta.2020.05.003] [Medline: 32571528]
4. Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Schubert J, Bania J, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. J Travel Med 2020 May 18;27(3):taaa037 [FREE Full text] [doi: 10.1093/jtm/taaa037] [Medline: 32181488]
5. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Imperial College COVID-19 Response Team, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020 Aug;584(7820):257-261. [doi: 10.1038/s41586-020-2405-7] [Medline: 32512579]
6. Dye C, Cheng R, Dagpunar J, Williams B. The scale and dynamics of COVID-19 epidemics across Europe. R Soc Open Sci 2020 Nov;7(11):201726 [FREE Full text] [doi: 10.1098/rsos.201726] [Medline: 33391818]
7. Salje H, Tran Kiem C, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the burden of SARS-CoV-2 in France. Science 2020 Jul 10;369(6500):208-211 [FREE Full text] [doi: 10.1126/science.abc3517] [Medline: 32404476]
8. Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the effectiveness of government interventions against COVID-19. Science 2021 Feb 19;371(6531):eabd9338 [FREE Full text] [doi: 10.1126/science.abd9338] [Medline: 33323424]
9. Okell LC, Verity R, Watson OJ, Mishra S, Walker P, Whittaker C, et al. Have deaths from COVID-19 in Europe plateaued due to herd immunity? Lancet 2020 Jun 20;395(10241):e110-e111 [FREE Full text] [doi: 10.1016/S0140-6736(20)31357-X] [Medline: 32534627]
10. Sibley CG, Greaves LM, Satherley N, Wilson MS, Overall NC, Lee CHJ, et al. Effects of the COVID-19 pandemic and nationwide lockdown on trust, attitudes toward government, and well-being. Am Psychol 2020;75(5):618-630. [doi: 10.1037/amp0000662] [Medline: 32496074]
11. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: Digital topic modeling approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]
12. Pan SL, Cui M, Qian J. Information resource orchestration during the COVID-19 pandemic: A study of community lockdowns in China. Int J Inf Manage 2020 Oct;54:102143 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102143] [Medline: 32394997]
13. 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):2153-2157. [doi: 10.1056/nejmp0900702]
14. Freifeld C, Mandl K, Reis B, Brownstein J. HealthMap: Global infectious disease monitoring through automated classification and visualization of internet media reports. J Am Med Inform Assoc 2008;15(2):150-157 [FREE Full text] [doi: 10.1197/jamia.M2544] [Medline: 18096908]
15. Hamzah FAB, Lau CH, Nazri H, Ligot DV, Lee G, Tan CL, et al. CoronaTracker: World-wide COVID-19 outbreak data analysis and prediction. Bull World Health Organ 2020 Mar 19:1 [FREE Full text] [doi: 10.2471/blt.20.255695]
16. Evanega S, Lynas M, Adams J, Smolenyak K. Coronavirus misinformation: Quantifying sources and themes in the COVID-19 'infodemic'. JMIR Preprints. Preprint posted online on October 19, 2020. [FREE Full text] [doi: 10.2196/preprints.25143]
17. Madraki G, Grasso I, Otala J, Liu Y, Matthews J. Characterizing and comparing COVID-19 misinformation across languages, countries and platforms. arXiv. Preprint posted online on October 13, 2020. [FREE Full text]
18. Okan O, Bollweg TM, Berens E, Hurrelmann K, Bauer U, Schaeffer D. Coronavirus-related health literacy: A cross-sectional study in adults during the COVID-19 infodemic in Germany. Int J Environ Res Public Health 2020 Jul 30;17(15):5503 [FREE Full text] [doi: 10.3390/ijerph17155503] [Medline: 32751484]
19. Tavares Lima CK, Moreira de Medeiros Carvalho P, de Araújo Araruna Silva Lima I, de Oliveira Nunes JVA, Saraiva JS, de Souza RI, et al. The emotional impact of coronavirus 2019-nCoV (new coronavirus disease). Psychiatry Res 2020 May;287:112915 [FREE Full text] [doi: 10.1016/j.psychres.2020.112915] [Medline: 32199182]
20. Shabahang R, Aruguete M, McCutcheon L. Online health information utilization and online news exposure as predictor of COVID-19 anxiety. N Am J Psychol 2020 Sep 28;22(3):469-482 [FREE Full text]
21. Hamidein Z, Hatami J, Rezapour T. How people emotionally respond to the news on COVID-19: An online survey. Basic Clin Neurosci 2020;11(2):171-178 [FREE Full text] [doi: 10.32598/bcn.11.covid19.809.2] [Medline: 32855776]
22. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, et al. Social network analysis of COVID-19 sentiments: Application of artificial intelligence. J Med Internet Res 2020 Aug 18;22(8):e22590 [FREE Full text] [doi: 10.2196/22590] [Medline: 32750001]
23. Yin H, Yang S, Li J. Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media. In: Proceedings of the International Conference on Advanced Data Mining and Applications. 2020 Presented at: International Conference

on Advanced Data Mining and Applications; November 12-15, 2020; Foshan, China p. 610-623 URL: https://arxiv.org/pdf/2007.02304 [doi: 10.1007/978-3-030-65390-3_46]

24. Kumar A, Khan SU, Kalra A. COVID-19 pandemic: A sentiment analysis. Eur Heart J 2020 Oct 14;41(39):3782-3783 [FREE Full text] [doi: 10.1093/eurheartj/ehaa597] [Medline: 32678890]
25. Aslam F, Awan TM, Syed JH, Kashif A, Parveen M. Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. Humanit Soc Sci Commun 2020 Jul 08;7(1):1-9 [FREE Full text] [doi: 10.1057/s41599-020-0523-3]
26. Dong M, Zheng J. Letter to the editor: Headline stress disorder caused by Netnews during the outbreak of COVID-19. Health Expect 2020 Apr;23(2):259-260 [FREE Full text] [doi: 10.1111/hex.13055] [Medline: 32227627]
27. Chakraborty A, Bose S. Around the world in 60 days: An exploratory study of impact of COVID-19 on online global news sentiment. J Comput Soc Sci 2020 Oct 21:1-34 [FREE Full text] [doi: 10.1007/s42001-020-00088-3] [Medline: 33102926]
28. Krawczyk K, Volanakis A. SciRide. URL: http://sciride.org [accessed 2021-05-09]
29. Flaxman S, Goel S, Rao JM. Filter bubbles, echo chambers, and online news consumption. Public Opin Q 2016 Mar 22;80(S1):298-320. [doi: 10.1093/poq/nfw006]
30. Internet Archive. URL: http://web.archive.org/ [accessed 2021-05-23]
31. Volanakis A, Krawczyk K. SciRide Finder: A citation-based paradigm in biomedical literature search. Sci Rep 2018 Apr 18;8(1):6193 [FREE Full text] [doi: 10.1038/s41598-018-24571-0] [Medline: 29670147]
32. Trampuš M, Novak B. Internals of an aggregated web news feed. In: Proceedings of the 15th International Multiconference Information Society. 2012 Presented at: 15th International Multiconference Information Society; October 8-12, 2012; Ljubljana, Slovenia p. 221-224 URL: http://library.ijs.si/Stacks/Proceedings/InformationSociety/2012/IS2012_Volume_A.pdf
33. Gilbert CJ, Hutto E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. 2014 Presented at: 8th International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI p. 216-225 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109/8122
34. Agarwal A. Sentiment analysis of financial news. In: Proceedings of the 12th International Conference on Computational Intelligence and Communication Networks. 2020 Presented at: 12th International Conference on Computational Intelligence and Communication Networks; September 25-26, 2020; Bhimtal, India p. 312-315. [doi: 10.1109/cicn49253.2020.9242579]
35. Lemire M, Paré G, Sicotte C, Harvey C. Determinants of internet use as a preferred source of information on personal health. Int J Med Inform 2008 Nov;77(11):723-734. [doi: 10.1016/j.ijmedinf.2008.03.002] [Medline: 18434246]
36. Gledson A, Keane J. Measuring topic homogeneity and its application to dictionary-based word sense disambiguation. In: Proceedings of the 22nd International Conference on Computational Linguistics. 2008 Presented at: 22nd International Conference on Computational Linguistics; August 18-22, 2008; Manchester, UK p. 273-280. [doi: 10.3115/1599081.1599116]
37. Jobes D, Berman A, O'Carroll P, Eastgard S, Knickmeyer S. The Kurt Cobain suicide crisis: Perspectives from research, public health, and the news media. Suicide Life Threat Behav 1996;26(3):260-269; discussion 269-271. [Medline: 8897665]
38. Brown W, Basil M. Media celebrities and public health: Responses to 'Magic' Johnson's HIV disclosure and its impact on AIDS risk and high-risk behaviors. Health Commun 1995 Oct;7(4):345-370. [doi: 10.1207/s15327027hc0704_4]
39. Berry TR, Wharf-Higgins J, Naylor PJ. SARS wars: An examination of the quantity and construction of health information in the news media. Health Commun 2007;21(1):35-44. [doi: 10.1080/10410230701283322] [Medline: 17461750]
40. Weitz JS, Park SW, Eksin C, Dushoff J. Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations. Proc Natl Acad Sci U S A 2020 Dec 22;117(51):32764-32771 [FREE Full text] [doi: 10.1073/pnas.2009911117] [Medline: 33262277]

Abbreviations

GDELT: Global Database of Events, Language, and Tone
NIHR: National Institute for Health Research
NPI: nonpharmaceutical intervention
ONS: online news source
RSS: Really Simple Syndication
***rsskew*:** relative sentiment skew

Edited by C Basch; submitted 28.02.21; peer-reviewed by J Huang, B Chen; comments to author 01.04.21; revised version received 18.04.21; accepted 18.04.21; published 02.06.21.

Please cite as:
Krawczyk K, Chelkowski T, Laydon DJ, Mishra S, Xifara D, Gibert B, Flaxman S, Mellan T, Schwämmle V, Röttger R, Hadsund JT, Bhatt S
Quantifying Online News Media Coverage of the COVID-19 Pandemic: Text Mining Study and Resource
J Med Internet Res 2021;23(6):e28253
URL: https://www.jmir.org/2021/6/e28253
doi:10.2196/28253
PMID:33900934

Original Paper

Partisan Differences in Twitter Language Among US Legislators During the COVID-19 Pandemic: Cross-sectional Study

Sharath Chandra Guntuku[1,2,3], PhD; Jonathan Purtle[4], DrPH; Zachary F Meisel[3,5,6], MPH, MD; Raina M Merchant[2,3,5,6], MD; Anish Agarwal[2,3,5,6], MPH, MS, MD

[1]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, United States
[2]Penn Medicine Center for Digital Health, University of Pennsylvania, Philadelphia, PA, United States
[3]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States
[4]Department of Health Management & Policy, Drexel University Dornsife School of Public Health, Philadelphia, PA, United States
[5]Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA, United States
[6]Center for Emergency Care Research and Policy, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Sharath Chandra Guntuku, PhD
Department of Computer and Information Science
University of Pennsylvania
3300 Walnut St
Philadelphia, PA, 19103
United States
Phone: 1 215 615 3211
Email: sharathg@cis.upenn.edu

Abstract

Background: As policy makers continue to shape the national and local responses to the COVID-19 pandemic, the information they choose to share and how they frame their content provide key insights into the public and health care systems.

Objective: We examined the language used by the members of the US House and Senate during the first 10 months of the COVID-19 pandemic and measured content and sentiment based on the tweets that they shared.

Methods: We used Quorum (Quorum Analytics Inc) to access more than 300,000 tweets posted by US legislators from January 1 to October 10, 2020. We used differential language analyses to compare the content and sentiment of tweets posted by legislators based on their party affiliation.

Results: We found that health care–related themes in Democratic legislators' tweets focused on racial disparities in care (odds ratio [OR] 2.24, 95% CI 2.22-2.27; *P*<.001), health care and insurance (OR 1.74, 95% CI 1.7-1.77; *P*<.001), COVID-19 testing (OR 1.15, 95% CI 1.12-1.19; *P*<.001), and public health guidelines (OR 1.25, 95% CI 1.22-1.29; *P*<.001). The dominant themes in the Republican legislators' discourse included vaccine development (OR 1.51, 95% CI 1.47-1.55; *P*<.001) and hospital resources and equipment (OR 1.22, 95% CI 1.18-1.25). Nonhealth care–related topics associated with a Democratic affiliation included protections for essential workers (OR 1.55, 95% CI 1.52-1.59), the 2020 election and voting (OR 1.31, 95% CI 1.27-1.35), unemployment and housing (OR 1.27, 95% CI 1.24-1.31), crime and racism (OR 1.22, 95% CI 1.18-1.26), public town halls (OR 1.2, 95% CI 1.16-1.23), the Trump Administration (OR 1.22, 95% CI 1.19-1.26), immigration (OR 1.16, 95% CI 1.12-1.19), and the loss of life (OR 1.38, 95% CI 1.35-1.42). The themes associated with the Republican affiliation included China (OR 1.89, 95% CI 1.85-1.92), small business assistance (OR 1.27, 95% CI 1.23-1.3), congressional relief bills (OR 1.23, 95% CI 1.2-1.27), press briefings (OR 1.22, 95% CI 1.19-1.26), and economic recovery (OR 1.2, 95% CI 1.16-1.23).

Conclusions: Divergent language use on social media corresponds to the partisan divide in the first several months of the course of the COVID-19 public health crisis.

(J Med Internet Res 2021;23(6):e27300) doi:10.2196/27300

KEYWORDS

Twitter; COVID-19; digital health; US legislators; natural language processing; policy makers; social media; policy; politics; language; cross-sectional; content; sentiment; infodemiology; infoveillance

Introduction

The novel COVID-19 pandemic continues to surge throughout the world. The United States' federal and state policy responses continue to shift and vary throughout the stages of the pandemic [1]. Notable divisions related to public health measures and frameworks for closing and reopening local economies have proliferated [2]. A unique aspect of the COVID-19 pandemic is the role that social media plays in housing, disseminating, and amplifying information and opinions [3,4]. US legislators have also taken to social media to connect with their constituents, comment on the pandemic, and provide information across a spectrum of pandemic-related content to individuals.

Understanding what content US legislators are sharing through social media posts (eg, Twitter) and how they are relaying COVID-19–related information is important, as these issues guide public knowledge and public opinion and inform policy change. By using social media data, prior studies have identified growing partisan differences among Republican and Democrat legislators as the pandemic has progressed [5]. It has also been found that tweets about specific topics (eg, social distancing) from legislators are often associated with the time when policies are put into action, and the effect of such tweets are larger in democratic counties [6].

The objective of this study was to analyze the language in posts on Twitter—a leading social media platform—that were posted by US legislators over the course of the pandemic to identify potential health care–related themes in COVID-19–related posts and to analyze the associated sentiment within tweet language across partisans.

Methods

Data

We identified state legislators' Twitter posts that were related to COVID-19 and posted from January 1 to October 10, 2020, by using Quorum (Quorum Analytics Inc) [7], a software platform that collects policy-related documents, including social media posts from politicians that were posted during their time in office. This study was considered exempt from review by the University of Pennsylvania Institutional Review Board, as it involves the analysis of public-facing data.

Language Feature Extraction

We extracted the relative frequency of single words and phrases from tweets by using the Differential Language Analysis ToolKit package [8] and created two sets of features—(1) an open vocabulary that was defined by using latent Dirichlet allocation [9], an unsupervised clustering algorithm, to create 50 data-driven word clusters (topics) and (2) sentiment, which was measured by using the National Research Council (NRC) Canada lexicon [10], a data-driven dictionary containing words associated with positive and negative valence. The NRC lexicon was developed by using a corpus of 77,500 positive and negative tweets, and consists of 54,129 weighted unigrams and 316,531 bigrams in which the weight corresponds to the degree of association between a token and sentiment [10].

Statistical Analyses

To distinguish linguistic differences across political parties (coded as a dichotomous outcome), each feature set was input in a logistic regression model, and those that were significantly different according to a cutoff Benjamini-Hochberg–corrected P value of <.001 were reported [11]. Two authors independently evaluated each topic for thematic meanings by reviewing the top 10 posts per topic and coded them into health care–related and nonhealth care–related themes.

Data on changes in the prevalence and sentiment of topics that were significantly associated with either party and occurred over time were obtained by calculating the mean scores across all posts per week, stratified by party, and visualized via locally estimated scatterplot smoothing [12].

Results

US Legislators' Tweets

We identified 309,438 COVID-19–related tweets from the 4224 unique accounts of US legislators. The descriptive statistics of the data set are in Table 1. The number of tweets per legislator over the selected time period is shown in Multimedia Appendix 1. Tweet language that correlated with US legislature party affiliation is displayed in Figure 1. Of the statistically significant topics, we identified 7 health care–related themes and 14 nonhealth care–related themes associated with the two major party affiliations.

Figure 1. Words and phrases that were significantly associated with tweets from Democratic legislators (blue) and Republican legislators (red). Bar length indicates effect size and shade indicates relative word frequency (*P*<.001; Benjamini-Hochberg p-correction). CCP: Chinese Communist Party.

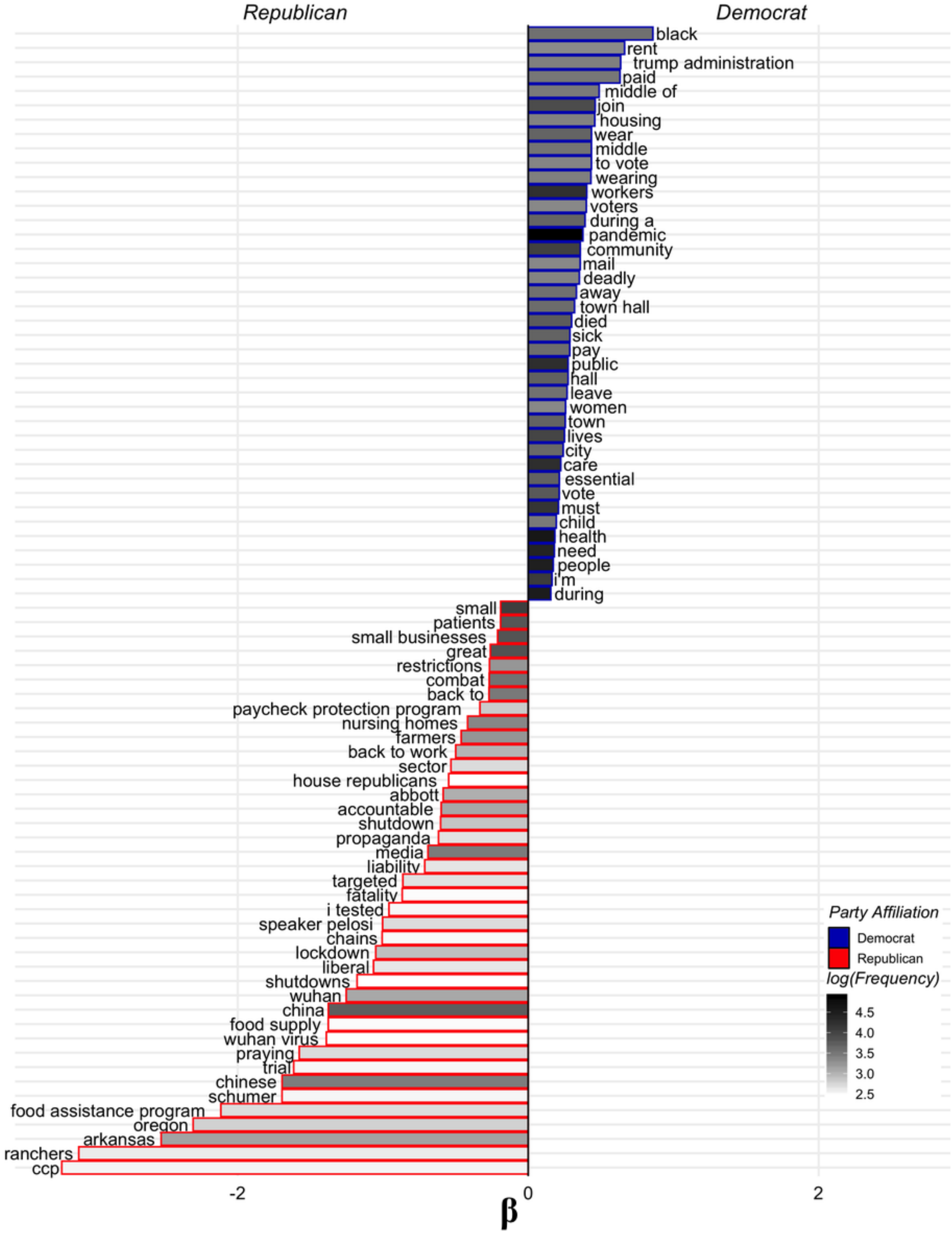


Table 1. Descriptive statistics of the data collected from Quorum from January 1 to October 10, 2020.

Category	Value, n (%)
Number of tweets	
All legislators	309,438 (100)
Republican legislators	88,146 (28.4)
Democrat legislators	221,292 (71.5)
Number of retweets	
Republican retweets	38,120 (12.3)
Democrat retweets	96,469 (31.1)
Number of individual accounts	
All legislators	4224 (100)
Democrat legislators	2432 (57.6)
Republican legislators	1792 (42.4)

Thematic Differences by Party Affiliation

Health care–related themes (Table 2) associated with a Democratic party affiliation included the following: racial disparities in care (odds ratio [OR] 2.24, 95% CI 2.22-2.27), health care and insurance (OR 1.74, 95% CI 1.70-1.77), public health guidelines (OR 1.25, 95% CI 1.22-1.29), and COVID-19 testing (OR, 1.15, 95% CI 1.12-1.19). Health care–related themes associated with a Republican party affiliation included the following: vaccine development (OR 1.51, 95% CI 1.47-1.55) and hospital resources and equipment (OR 1.22, 95% CI 1.18-1.25).

Nonhealth care–related topics were also identified across parities. The themes associated with a Democratic affiliation included the following: protections for essential workers, the 2020 election and voting, unemployment and housing, crime and racism, public town halls, the Trump Administration, immigration, and the loss of life. The themes associated with a Republican affiliation included the following: China, congressional relief bills, small business assistance, press briefings, and economic recovery (Table 3). The prevalence of the themes over time stratified by affiliation is shown in Multimedia Appendices 2 and 3. The set of topics that did not significantly correlate with affiliation are shown in Multimedia Appendix 4.

Table 2. Health care–related topics that are more likely to be posted by Democrat legislators and Republican legislators. Effect size is shown by using odds ratios (ORs) along with 95% CIs. Only significant topics after Benjamini-Hochberg p-correction (*P*<.001) are shown.

Affiliation and topic theme	Top words	ORs (95% CI)	Example tweets
Democrat			
Racial disparities	“communities,” “black,” “color,” “racial,” “disparities,” “disproportionately,” “impact,” “hit,” “women,” and “latino”	2.24 (2.22-2.27)	“COVID-19 is disproportionately harming communities of color and exposing generations of systemic racism. We need to collect racial and ethnic data for coronavirus testing and treatment so we can address these health disparities and begin rectifying decades of injustice.”
Health care and insurance	“healthcare,” “americans,” “access,” “insurance,” “coverage,” “affordable,” “millions,” “medicaid,” and “court”	1.74 (1.70-1.77)	“The #MedicareCrisisProgram would: Expand Medicare to the recently unemployed & cap out-of-pocket costs Expand Medicaid to cover even more people Ensure no out-of-pocket costs for COVID-19 testing/treatment for everyone”
Public health guidelines	“mask,” “social,” “distancing,” “wear,” “spread,” “hands,” “stay,” “home,” “wash,” “practice,” and “stop”	1.25 (1.22-1.29)	“The pandemic is nowhere near over. Continue practicing social distancing & WEAR A MASK”
COVID-19 testing	“testing,” “free,” “county,” “sites,” “appointment,” “center,” “residents,” “open,” “city,” “visit,” and “symptoms”	1.15 (1.12-1.19)	“New COVID-19 mobile testing site opens. Scheduled Locations for Free Drive-Through COVID-19 Testing” [retweet]
Republican			
Vaccine development	“vaccine,” “research,” “drug,” “production,” “effective,” “vaccines,” “treatments,” “development,” and “defense”	1.51 (1.47-1.55)	“Three Coronavirus Vaccine Developers Report Promising Initial Results” [retweet]
Hospital resources and equipment	“medical,” “patients,” “hospitals,” “masks,” “equipment,” “supplies,” “ppe,”[a] “donate,” “blood,” “needed,” and “plasma”	1.22 (1.18-1.25)	“Kansans everywhere are stepping up to fight the #Coronavirus.…,which manufactures aircraft parts in…, is using their 3D printing capabilities to work with local area hospitals on prototypes of N-95 surgical masks and protective face shields.”

[a] PPE: personal protective equipment.

Table 3. Nonhealth care–related topics that are more likely to be posted by Democratic legislators and Republican legislators. Effect size is shown by using odds ratios (ORs) along with 95% CIs. Only significant topics after Benjamini-Hochberg p-correction (*P*<.001) are shown.

Affiliation and topic theme	Top words	OR (95% CI)	Example tweets
Democrat			
Protections for essential workforce	“workers,” “essential,” “leave,” “sick,” “pay,” “employees,” “safety,” “job,” “frontline,” “grocery,” “protections,” and “deserve”	1.55 (1.52-1.59)	“Essential workers--like farmworkers, first responders, health care workers, and grocery store workers--deserve hazard pay from their government or their company for their service during the #coronavirus pandemic.”
2020 election and voting	“vote,” “mail,” “voters,” “ballot,” “elections,” “absentee,” “november,” “census,” “primary,” “voter,” “ballots,” and “2020”	1.31 (1.27-1.35)	“To make sure this virus doesn't keep people from the ballot box, states and localities should bring the ballot box to them through expanded vote-by-mail and no-fault absentee voting. #SAFEDemocracy”
Unemployment and housing assistance	“unemployment,” “assistance,” “benefits,” “rent,” “program,” “housing,” “insurance,” “eviction,” “eligible,” “lost,” and “claims”	1.27 (1.24-1.31)	“Wisconsin residents who have exhausted their regular unemployment insurance (UI) benefits may now apply for Pandemic Emergency Unemployment Compensation (PEUC).”
Crime and racism	“violence,” “police,” “domestic,” “racism,” “asian,” “stand,” “hate,” “gun,” “survivors,” “protests,” “discrimination,” and “victims”	1.22 (1.18-1.26)	“As coronavirus fears have intensified, incidents of violence & discrimination against Chinese Americans have increased. Joined hundreds of San Franciscans marching in Chinatown today to protest prejudice and racial profiling. #TogetherWeStand #StandWithChinatown”
Public town halls	“join,” “hall,” “town,” “live,” “questions,” “discuss,” “tomorrow,” “tune,” “facebook,” “virtual,” “i'll,” “telephone,” “tonight”	1.2 (1.16-1.23)	“Tomorrow evening, at 7:30 pm EST [Eastern Standard Time], I am hosting another Coronavirus Telephone Town Hall.”
Trump Administration			
Word set 1	“trump,” “national,” “service,” “global,” “administration,” “postal,” “guard,” “security,” “decision,” “usps,”[a] and “members”	1.22 (1.19-1.26)	“Donald Trump and Mike Pence's handling of COVID-19 is the greatest failure of any American presidency.…#Debate2020 #TrumpFailure #IwillVote”
Word set 2	“trump,” “president,” “white,” “house,” “trump's,” “donald,” “don't,” “biden,” “he's,” “joe,” “administration,” “campaign,” “pence,” “force,” and “rally”	1.17 (1.13-1.2)	“Democratic presidential candidate Joe Biden criticized President Donald Trump's 'callousness' in handling the coronavirus pandemic” [retweet]
Immigration	“letter,” “release,” “colleagues,” “risk,” “prisons,” “urging,” “ice,”[b] “immigration,” “inmates,” “detention,” and “vulnerable”	1.16 (1.12-1.19)	“ICE must suspend immigration enforcement during the #COVID19 pandemic.”
Loss of life			
Word set 1	“trump,” “americans,” “lives,” “president,” “american,” “died,” “100,” “million,” “deaths,” “200,” “states,” “lost,” “dead,” and “leadership”	1.38 (1.35-1.42)	“In trump's catastrophic zeal to lie about the coronavirus threat, innumerable Americans' lives are in danger.”
Word set 2	“family,” “lost,” “friends,” “loved,” “died,” “life,” “heart,” “loss,” “years,” “god,” “prayers,” “remember,” “thoughts,” and “honor”	1.13 (1.09-1.16)	“Within a few short days, 200,000 mothers fathers daughters sons children parents lovers wives husbands friends grandparents aunts uncles cousins DEAD of coronavirus. We dare not get numb. That equates to every person...EVERY single person in my community.”
Republican			
China	“china,” “world,” “chinese,” “accountable,” “communist,” “global,” “hold,” “party,” “wuhan,” “china's,” “outbreak,” “government,” “travel,” “held,” and “organization”	1.89 (1.85-1.92)	“Pompeo slams communist China for lying about Wuhan coronavirus” and “The Wuhan virus is #MadeInChina.” [retweet]
Small business assistance	“small,” “businesses,” “program,” “relief,” “loans,” “impacted,” “apply,” “owners,” “economic,” “grants,” “assistance,” “grant,” and “disaster”	1.27 (1.23-1.3)	“Small businesses impacted by the #coronavirus can apply for a low-interest SBA [Small Business Association] disaster loan here”

Affiliation and topic theme	Top words	OR (95% CI)	Example tweets
Government relief funds			
Word set 1	“funding,” “federal,” “local,” “state,” “million,” “act,” “support,” “relief,” “communities,” “cares,” “resources,” “governments,” “provide,” and “emergency”	1.22 (1.18-1.25)	“Democrats Filibuster Covid Relief” [retweet]
Word set 2	“relief,” “senate,” “bill,” “republicans,” “democrats,” “house,” “americans,” “mcconnell,” “pass,” “american,” “pelosi,” “congress,” “package,” “politics,” and “gop”[c]	1.23 (1.2-1.27)	“In the CARES [Coronavirus Aid, Relief, and Economic Security] Act,…received money to be disbursed ASAP [as soon as possible] to help local communities fight coronavirus. Grants were disbursed through specific programs, including Community Development Block Grants-flexible funding to states and local governments-Steubenville awarded $365,667!”
Press briefings			
Word set 1	“governor,” “update,” “press,” “north,” “conference,” “gov,” “watch,” “live,” “carolina,” “#ncpol,”[d] “michigan,” “nc,”[e] “south,” “state's,” and “briefing”	1.22 (1.19-1.26)	“What is Montana doing in response to coronavirus?” COVID-19 Montana state response update.”
Word set 2	“hearing,” “response,” “discuss,” “committee,” “watch,” “force,” “joined,” “task,” “meeting,” “dr,”[f] “morning,” “impact,” “talk,” “yesterday,” and “hear”	1.14 (1.1-1.17)	“#ICYMI [in case you missed it] I discussed the #Coronavirus in depth on the latest episode of #TheBreakDown.”
State politics	“georgia,” “court,” “power,” “supreme,” “restrictions,” “republican,” “politics,” “tennessee,” “decision,” “law,” “abortion,” and “wisconsin”	1.16 (1.12-1.19)	“6th Circuit unanimously rules DWP (drunk-with-power) Beshear ban on church services unconstitutional! Kentucky coronavirus”
Economic recovery and news			
Word set 1	“economy,” “economic,” “back,” “jobs,” “recovery,” “plan,” “recover,” “america,” “future,” “nation,” “forward,” “industry,” “climate,” “energy,” and “safely”	1.19 (1.15-1.22)	“As a nation, we will defeat the coronavirus and rebuild the greatest economy. #CommitmentToAmerica”
Word set 2	“news,” “times,” “york,” “data,” “good,” “washington,” “thread,” “study,” “post,” “breaking,” “shows,” “found,” and “months”	1.20 (1.16-1.23)	“Germany's R0 coronavirus experiment: Berlin tries to manage a variable no one can measure accurately”

[a] USPS: US Postal Service.

[b] ICE: Immigration and Customs Enforcement.

[c] GOP: Grand Old Party.

[d] NCPOL: North Carolina Political News.

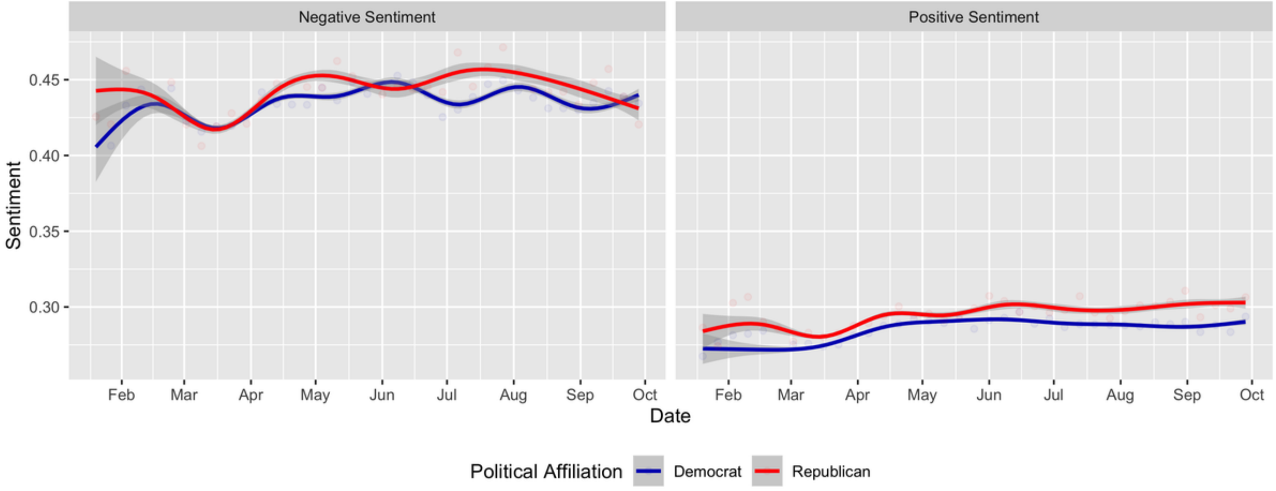
[e] NC: North Carolina.

[f] DR: doctor.

Sentiment Differences by Party Affiliation

We performed an analysis of sentiment for the language used in tweets and found that overall, Republican-affiliated tweets used more positive sentiment, which increased over time. The variation in overall sentiment is shown in Figure 2. Negative sentiment was associated with content from both parties across the following themes: health care and insurance, COVID-19 testing, and racial disparities. Positive sentiment was associated with content within the theme of government public health expertise. Sentiment within themes over time and across parities is identified in Multimedia Appendices 5 and 6.

Figure 2. Sentiment analysis of US legislators' language on Twitter across party affiliations.



Discussion

By using machine learning techniques, we investigated narrative content in over 300,000 twitter posts from US legislators over the course of the COVID-19 pandemic to date. Investigating the language within posts on social media platforms has become more common and has been specifically used to study aspects of health and health care. This study is among the first to analyze US legislators' Twitter-based language to identify the COVID-19–related themes that policy makers are discussing on Twitter with a specific focus on health care–related topics. Additionally, this study deployed advanced language assessments that use machine learning to analyze how legislators are talking about these themes by conducting sentiment analyses throughout the phases of the pandemic.

We noted key differences across the two major US political parties. Health care–related themes that correlated with a Democratic party affiliation focused on the health care access and disparities across race. The themes that correlated with a Republican party affiliation focused on initial and persistent vaccine progress, access to equipment (eg, personal protective equipment), and government expertise. Furthermore, in the language analysis, we identified that across content posted by Republican legislators, there was considerably more content about the pandemic and approaches for managing the pandemic across health care topics. Language analysis was also used to detect thematic differences in narrative content within Twitter posts across the two major political parties. In this study, our results indicated that legislators with a Democratic party affiliation focused their COVID-19 content more toward social services and racial disparities. Content from Republican-affiliated legislators focused thematically on government relief and economic aid. This finding is consistent with surveys of elected officials and the general public, which suggests that awareness and concern about health disparities among Democrats are greater than those among Republicans [13,14].

There are limitations to this study, including the fact that content was collected from publicly available Twitter posts; thus, legislators who do not post content were not included. If a legislator did not have a party affiliation (as noted by the Quorum database), we could not include them in this analysis. We also did not control for demographic or health access data, as our analysis was performed on the language of individual legislators. Further, a topic's significant association with a particular affiliation does not imply that other party legislators did not tweet about it; it only indicates the relative frequency of tweets containing the words that were associated with each topic.

This study highlights the ability to understand how legislators use social media (eg, Twitter); what information they choose to share; and how they frame their content, which was determined through sentiment analysis [15]. These are key insights that will remain important to the public and health care systems as policy makers continue to shape the national and local responses to the pandemic [16].

Acknowledgments

The authors thank Garrick Sherman for inputs on visualizations. SCG received support from Google Cloud. Support for this study, including data acquisition, was provided by the Center for Health Economics of Treatment Interventions for Substance Use Disorder, HCV (hepatitis C virus), and HIV and a National Institute on Drug Abuse research center (R21DA050761-02). This study was conducted in partnership with the Research-to-Policy Collaboration, which is affiliated with The Pennsylvania State University's Edna Bennett Pierce Prevention Research Center.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mean number of COVID-19–related tweets per legislator by party affiliation. The tweets in our data set were posted from January 1 to October 10, 2020.
[PNG File , 140 KB - jmir_v23i6e27300_app1.png]

Multimedia Appendix 2
Prevalence of health care–related topics over time by party affiliation.
[PNG File , 277 KB - jmir_v23i6e27300_app2.png]

Multimedia Appendix 3
Prevalence of nonhealth care–related topics over time by party affiliation.
[PNG File , 286 KB - jmir_v23i6e27300_app3.png]

Multimedia Appendix 4
Topics that were associated with Democrat (top) and Republican (bottom) legislators' Twitter language but did not pass the statistical significance threshold.
[DOCX File , 15 KB - jmir_v23i6e27300_app4.docx]

Multimedia Appendix 5
Sentiment of COVID-19–related legislator tweets on health care topics.
[PNG File , 281 KB - jmir_v23i6e27300_app5.png]

Multimedia Appendix 6
Sentiment of COVID-19–related legislator tweets on nonhealth care topics.
[PNG File , 304 KB - jmir_v23i6e27300_app6.png]

References

1. Gadarian SK, Goodman SW, Pepinsky TB. Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 Pandemic. SSRN. Preprint posted online on March 27, 2020. [doi: 10.2139/ssrn.3562796]
2. Bromfield N, McConnell A. Two routes to precarious success: Australia, New Zealand, COVID-19 and the politics of crisis governance. International Review of Administrative Sciences 2020 Dec 01:1-18 [FREE Full text] [doi: 10.1177/0020852320972465]
3. Mheidly N, Fares J. Leveraging media and health communication strategies to overcome the COVID-19 infodemic. J Public Health Policy 2020 Dec;41(4):410-420 [FREE Full text] [doi: 10.1057/s41271-020-00247-w] [Medline: 32826935]
4. Ojo A, Guntuku SC, Zheng M, Beidas RS, Ranney ML. How health care workers wield influence through Twitter hashtags: Retrospective cross-sectional study of the gun violence and COVID-19 public health crises. JMIR Public Health Surveill 2021 Jan 06;7(1):e24562 [FREE Full text] [doi: 10.2196/24562] [Medline: 33315578]
5. Green J, Edgerton J, Naftel D, Shoub K, Cranmer SJ. Elusive consensus: Polarization in elite communication on the COVID-19 pandemic. Sci Adv 2020 Jul 10;6(28):eabc2717 [FREE Full text] [doi: 10.1126/sciadv.abc2717] [Medline: 32923600]
6. Grossman G, Kim S, Rexer J, Thirumurthy H. Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. SSRN. Preprint posted online on April 17, 2020. [doi: 10.2139/ssrn.3578695]
7. Best-in-class public affairs software. Quorum. URL: https://www.quorum.us/ [accessed 2021-05-10]
8. Schwartz HA, Giorgi S, Sap M, Crutchley P, Ungar L, Eichstaedt J. DLATK: Differential Language Analysis ToolKit. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2017 Sep Presented at: 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; September 2017; Copenhagen, Denmark p. 55-60. [doi: 10.18653/v1/d17-2010]
9. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003 Mar:993-1022 [FREE Full text] [doi: 10.5555/944919.944937]
10. Mohammad SM, Turney PD. Crowdsourcing a word–emotion association lexicon. Comput Intell 2012 Sep 04;29(3):436-465. [doi: 10.1111/j.1467-8640.2012.00460.x]
11. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodo 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
12. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 1979;74(368):829-836. [doi: 10.1080/01621459.1979.10481038]

13. Purtle J, Henson RM, Carroll-Scott A, Kolker J, Joshi R, Roux AVD. US mayors' and health commissioners' opinions about health disparities in their cities. Am J Public Health 2018 May;108(5):634-641. [doi: 10.2105/AJPH.2017.304298] [Medline: 29565663]
14. Gollust SE, Vogel RI, Rothman A, Yzer M, Fowler EF, Nagler RH. Americans' perceptions of disparities in COVID-19 mortality: Results from a nationally-representative survey. Prev Med 2020 Dec;141:106278 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106278] [Medline: 33027615]
15. Hansen MA, Johansson I, Sadowski K, Blaszczynski J, Meyer S. The partisan impact on local government dissemination of COVID-19 information: Assessing US county government websites. Can J Polit Sci 2020 Oct 12;54(1):150-162 [FREE Full text] [doi: 10.1017/s0008423920000918]
16. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]

Abbreviations

HCV: hepatitis C virus
NRC: National Research Council
OR: odds ratio

Edited by C Basch; submitted 20.01.21; peer-reviewed by D Huang, A Weichselbraun; comments to author 19.02.21; revised version received 25.02.21; accepted 16.04.21; published 03.06.21.

Please cite as:
Guntuku SC, Purtle J, Meisel ZF, Merchant RM, Agarwal A
Partisan Differences in Twitter Language Among US Legislators During the COVID-19 Pandemic: Cross-sectional Study
J Med Internet Res 2021;23(6):e27300
URL: https://www.jmir.org/2021/6/e27300
doi:10.2196/27300
PMID:33939620

Original Paper

Mental Health and Behavior of College Students During the COVID-19 Pandemic: Longitudinal Mobile Smartphone and Ecological Momentary Assessment Study, Part II

Dante L Mack[1], BA; Alex W DaSilva[1], BSc; Courtney Rogers[1], MEd; Elin Hedlund[1], BA; Eilis I Murphy[1], BA; Vlado Vojdanovski[2]; Jane Plomp[1], BA; Weichen Wang[2], ME; Subigya K Nepal[2], BSc; Paul E Holtzheimer[3,4], MD, MS; Dylan D Wagner[5], PhD; Nicholas C Jacobson[6], PhD; Meghan L Meyer[1], PhD; Andrew T Campbell[2], PhD; Jeremy F Huckins[1], PhD

[1]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, United States
[2]Department of Computer Science, Dartmouth College, Hanover, NH, United States
[3]National Center for PTSD, White River Junction, VT, United States
[4]Department of Psychiatry, Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States
[5]Department of Psychology, Ohio State University, Columbus, OH, United States
[6]Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States

Corresponding Author:
Dante L Mack, BA
Department of Psychological and Brain Sciences
Dartmouth College
Moore Hall
3 Maynard St
Hanover, NH, 03755
United States
Phone: 1 603 646 3181
Email: f002vhk@dartmouth.edu

Related Article:

Companion article: https://www.jmir.org/2020/6/e20185

Abstract

Background: Since late 2019, the lives of people across the globe have been disrupted by COVID-19. Millions of people have become infected with the disease, while billions of people have been continually asked or required by local and national governments to change their behavioral patterns. Previous research on the COVID-19 pandemic suggests that it is associated with large-scale behavioral and mental health changes; however, few studies have been able to track these changes with frequent, near real-time sampling or compare these changes to previous years of data for the same individuals.

Objective: By combining mobile phone sensing and self-reported mental health data in a cohort of college-aged students enrolled in a longitudinal study, we seek to understand the behavioral and mental health impacts associated with the COVID-19 pandemic, measured by interest across the United States in the search terms *coronavirus* and *COVID fatigue*.

Methods: Behaviors such as the number of locations visited, distance traveled, duration of phone use, number of phone unlocks, sleep duration, and sedentary time were measured using the StudentLife mobile smartphone sensing app. Depression and anxiety were assessed using weekly self-reported ecological momentary assessments, including the Patient Health Questionnaire-4. The participants were 217 undergraduate students. Differences in behaviors and self-reported mental health collected during the Spring 2020 term, as compared to previous terms in the same cohort, were modeled using mixed linear models.

Results: Linear mixed models demonstrated differences in phone use, sleep, sedentary time and number of locations visited associated with the COVID-19 pandemic. In further models, these behaviors were strongly associated with increased interest in *COVID fatigue*. When mental health metrics (eg, depression and anxiety) were added to the previous measures (week of term, number of locations visited, phone use, sedentary time), both anxiety and depression (*P*<.001) were significantly associated with interest in *COVID fatigue*. Notably, these behavioral and mental health changes are consistent with those observed around the initial implementation of COVID-19 lockdowns in the spring of 2020.

Conclusions: In the initial lockdown phase of the COVID-19 pandemic, people spent more time on their phones, were more sedentary, visited fewer locations, and exhibited increased symptoms of anxiety and depression. As the pandemic persisted through the spring, people continued to exhibit very similar changes in both mental health and behaviors. Although these large-scale shifts in mental health and behaviors are unsurprising, understanding them is critical in disrupting the negative consequences to mental health during the ongoing pandemic.

(J Med Internet Res 2021;23(6):e28892) doi:10.2196/28892

KEYWORDS

anxiety; college; COVID-19; COVID fatigue; depression; George Floyd; mobile sensing; phone usage; sleep; digital phenotyping

Introduction

The COVID-19 Pandemic

COVID-19 has drastically changed life for people worldwide. The World Health Organization (WHO) declared COVID-19 a pandemic on March 11, 2020 [1]. In this work, we extend our previous research beyond the initial phases of the pandemic to June 25, 2020. By this time, the world had seen over 2,000,000 confirmed COVID-19 cases and 100,000 deaths [2], and the severe long-term political, economic, educational, and social ramifications of COVID-19 had grown increasingly acute. Understanding the behavioral and mental health implications for individuals during the spring of 2020, an unprecedented period of high stress, is critical for the upkeep of mental health during the pandemic.

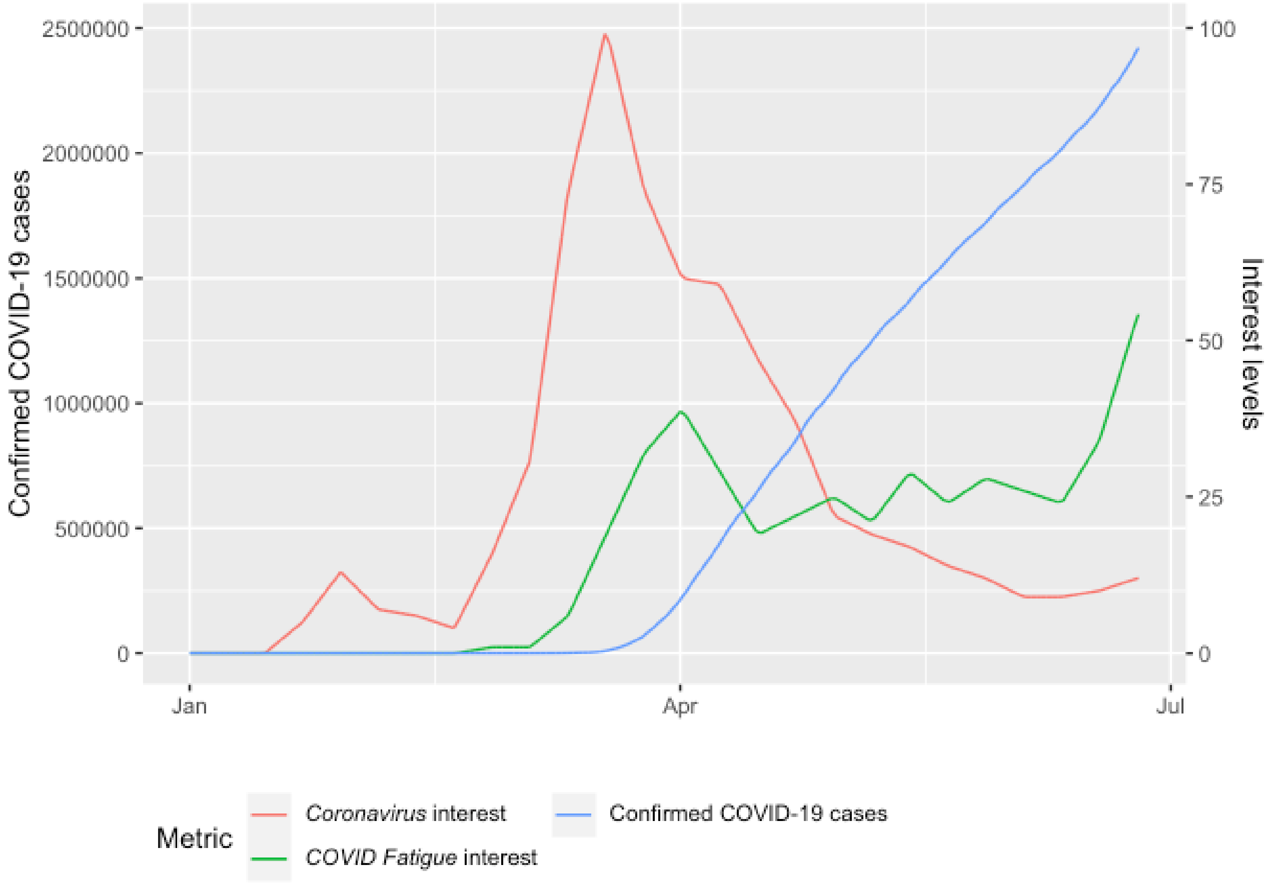
College students seem to be vulnerable to the impact of COVID-19 on mental health [3-5]. In the spring of 2020, global health measures forced many colleges to shut down in an effort to mitigate the spread of the virus. On March 12, 2020, Dartmouth College requested that students leave the campus, with hope that the campus could resume normal operations in 5 weeks. At that time, it was believed that the COVID-19 situation would end rather quickly; however, by March 17, the severity of the virus became more apparent, and Dartmouth College announced that all classes and office hours would be held remotely during the Spring 2020 academic term. Students were asked to leave campus and not return until further notice. Following changes in school, local, and federal policies, drastic alterations in students' behavior were observed. Prior research found that during the initial stages of the pandemic, this cohort of students visited fewer locations, spent more time sedentary, and increased its phone use [6]. All these behavioral changes were linked with increased levels of anxiety and depression [6], a trend observed by other researchers among various groups of college students [7,8]. Despite the negative impact on their mental health, many students continued to practice physical distancing with the hope of a swift return to normalcy. Unfortunately, improved knowledge about the ways in which COVID-19 spreads made it increasingly clear that physical distancing and other measures needed to continue for the sake of public health. Although some localities began to ease their restrictions and reopen in April 2020, it was still widely suggested that people avoid public gatherings and maintain physical distance from others.

COVID-19 Restrictions Over the Long Term

Towns and cities across the United States were forced to sustain COVID-19 "Stay Home, Stay Safe" restrictions in an effort to stem the spread of COVID-19. The effort of adhering to these restrictions seems to have bred fatigue among people nationwide. Although there is evidence that these stay-at-home-orders resulted in an immediate suppression of negative changes in mental health brought on during this time [9], people grew weary of the effort required to maintain constant vigilance. For many, staying at home began to feel like a monotonous rut—many referred to this feeling as "COVID fatigue." Search term interest in *COVID fatigue* is seen alongside interest in *coronavirus* and the total number of confirmed COVID-19 cases [2] (Figure 1). Interest in *COVID fatigue* began to grow about a month into the pandemic (spiking in April) and continued to grow throughout the spring, while interest in *coronavirus* peaked around the time when policy changes were implemented in March before sharply and steadily declining. Therefore, *COVID fatigue* became an appropriate metric for measuring pandemic intensity at a given moment in time, while interest in the search term *coronavirus* seemed to capture the initial pandemic-related changes, reaching peak interest levels in late March. because many students continued to live under "Stay Home, Stay Safe" restrictions throughout the spring, it is necessary to understand how the sustenance of these behaviors is related to mental health.

Over the last year, the COVID-19 pandemic has prompted multiple lockdowns, which have brought about increased interest in the search term *COVID fatigue*, a term that has quickly become the subject of much discussion in public health research [10-13]. Cross-sectional research in Istanbul, Turkey, suggests that those with higher education levels experience more COVID-19–based fatigue [14], which may be attributable to a greater likelihood of this population to adhere to virus protocols and practice physical distancing [14,15]. Recent work proposes that college students may experience increases in depression due to COVID fatigue [16]. Most work measuring COVID fatigue has been limited to the use of web-based surveys. Using smartphone mobile sensing for digital phenotyping enables increased sampling frequency [17] and an ecologically valid approach [18,19], particularly when paired with ecological momentary assessments (EMAs). How behavioral and mental health changes related to COVID fatigue among populations at high risk for mental health issues (eg, college students) remains an unresolved question that we investigate in this work.

Figure 1. Number of confirmed COVID-19 cases graphed alongside search term interest (scaled from 0-100) during the winter and spring of 2020. COVID-19 case data were taken from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [2]. Search term interest was taken from Google Trends.



Modeling Stages of the Pandemic With Changes in Behavior and Mental Health

Using previously successful methods, this work seeks to infer the behavioral and mental health impacts of the initial onset and duration of the COVID-19 pandemic among a cohort of Dartmouth College undergraduate students. Search query–related data are a promising metric in gathering population-based interest, which can be used to successfully identify changes in mental health [9,20,21]. Prior work has shown EMAs to be an excellent tool for gathering real-time assessments of mental health [22]. Notably, researchers have also used EMAs to assess mental health changes during the COVID-19 pandemic [6,23]. Interest in the search terms *coronavirus* and *COVID fatigue* was gathered from Google Trends to quantify the initial and ongoing phases of the pandemic (Figure 1). Interest levels were linearly inferred based on smartphone behavioral sensing data (gathered via passive sensing) and reported levels of anxiety and depression. Linear models are a useful tool in understanding the impact of the COVID-19 pandemic on daily behaviors and mental health in previous work, as a nearly identical method using COVID-19 media presence was previously successful [6]. Previously, we observed that the early media presence of COVID-19 was associated with increased levels of anxiety and depression as well as increased sedentary time and phone use [6]. This work seeks to build on prior work conducted by this and other groups to gain further insight into the ways in which COVID-19 has impacted daily life and mental health. We hypothesized that over nearly 3 months, students would continue to exhibit changes in depression, anxiety, and daily behaviors similar to those seen in the initial response to the COVID-19 pandemic.

Objective

In this work, we seek to understand the behaviors and mental health changes associated with the continuing phases of the pandemic, as inferred by interest in the search terms *COVID fatigue* and *coronavirus*.

Methods

Study Design

All data from this study were obtained from the StudentLife study [24], which is a longitudinal multimodal study designed to follow the experiences of undergraduate students throughout their academic tenure with a focus on mental health. Study components include smartphone mobile sensing through the StudentLife app [25], EMAs, and surveys focusing on a variety of college experience components and functional neuroimaging [26].

Participants

Data were collected from 219 participants who agreed to provide mobile sensing data via the StudentLife app [25]. One participant was removed from the study for having a phone incompatible with the app, and one participant withdrew within a week of starting the study. Data for both participants were excluded from further analyses. Of the remaining 217 participants, 67.8% (147) were female, with an age range of 18-22 years at the time of enrollment. Recruitment for this study began in August 2017 and concluded in November 2018. This study was reviewed and approved by Dartmouth College's Committee for the Protection of Human Subjects. All data analyzed were collected from September 9, 2017, through June 25, 2020.

Academic Terms

At Dartmouth College, the academic schedule consists of a flexible, year-round calendar that is roughly divided into 4 10-week academic terms (quarters), typically followed by 2 or more weeks of break. COVID-19 was first discovered in the United States during Dartmouth College's Winter 2020 term, which began on January 6, 2020. The Spring 2020 academic term (March 30 start date) was the first complete academic term that occurred during the COVID-19 pandemic. To graphically show differences in anxiety and depression related to COVID-19, academic terms, including the subsequent 2 weeks of break prior to the Winter 2020 term, were included as control terms.

Mobile Sensing and Ecological Momentary Assessments

Smartphone sensing data and EMA surveys were administered using the StudentLife application (iOS and Android) [25]. The StudentLife app collects data from several of the phone's sensors, including but not limited to GPS, accelerometer, and lock/unlock status. Anonymized data from the StudentLife app are uploaded to a secure server whenever a participant is both connected to a Wi-Fi network and charging their phone. Data from these sensors are used to assess items such as the day-to-day and week-by-week impact of workload on stress, sleep, activity, mood, sociability, mental well-being, and academic performance [25]. Students are prompted weekly by the StudentLife application to complete a few short surveys, administered as EMAs [22]. These EMAs include the Patient Health Questionnaire-4 (PHQ-4), which combines a brief measure of depressive and anxious symptoms [27] that assesses how often individuals were bothered by specific symptoms over the last 2 weeks with values ranging from 0 to 6 for each subscale. The PHQ-4 combines the Patient Health Questionnaire-2 (PHQ-2) and the Generalized Anxiety Disorder-2 (GAD-2) tool.

Sedentary Time

Sedentary time, or stationary duration, is computed with algorithms that detect a lack of movement based on accelerometer data from a phone's sensors. This enables us to measure students' activity or, more precisely, their lack of activity level. The StudentLife app continuously infers physical activities using the Android activity recognition application programming interface [28,29] or iOS Core Motion [30].

Sleep

Sleep was inferred through a combination of passive sensing features (ambient light, movement activity, screen on/off). In this way, three features were computed: sleep onset, wake time, and sleep duration. These measures of sleep have been shown to be accurate within ±30 minutes of total sleep duration [25].

Location

Density-based spatial clustering of applications with noise (DBSCAN) [31] was used to cluster GPS coordinates to determine the number of locations visited and distance traveled during a given time period. Locations were detected when 3 GPS samples (sampled every 10 minutes) were within a radius of 30 meters. Distance was calculated in meters traveled between all locations throughout the day.

Phone Use

Unlock duration is a measurement of time during which a user's phone is unlocked and the screen is on, calculated as the time between the user unlocking the phone and the user either manually relocking the phone or autolocking due to disuse (the iOS default is 30 seconds, Android defaults vary by manufacturer, and users can alter this time through their phone settings). Notification and system services do not influence the measurement of unlock duration. For the iOS app (189 users), from the start of the study in September 2017 until September 2018, unlock duration was measured by remotely triggering phones every 10 minutes, sampling 1 minute every 10 minute period (minimum 10% temporal coverage). After September 2018, phones were remotely triggered every 3 minutes, with subsequent sampling for 1 minute. Lock/unlock behaviors within that minute are recorded in real time, while locks/unlocks for the remaining 2 minutes are logged during the next remote trigger. The Android phone app natively supports phone use tracking and does not need to be triggered (28 users).

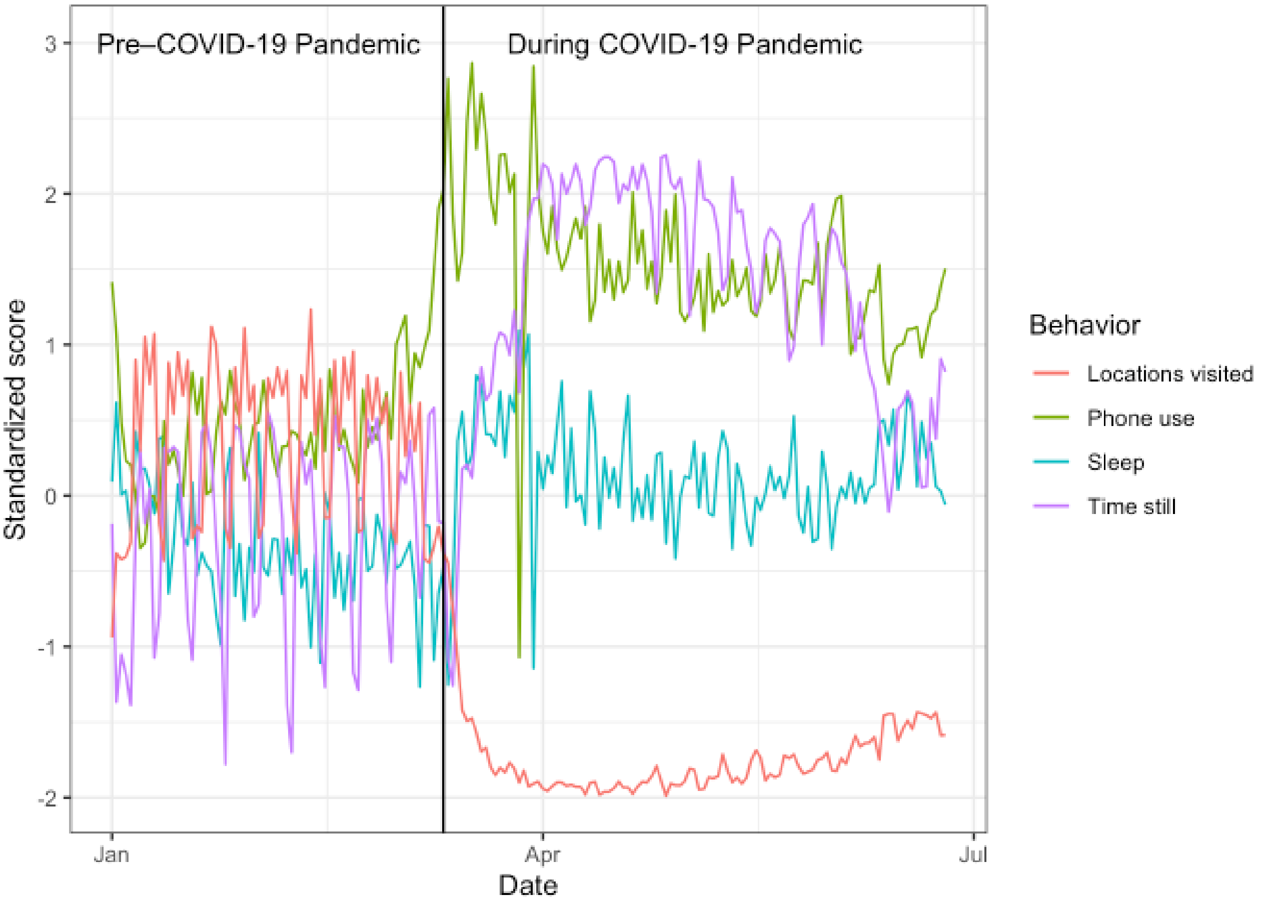
Google Trends

Google is the world's leading search engine, and Google Trends is an excellent tool to quantify topic interest. In turn, interest data from Google Trends have been successfully used to quantify changes in mental health [9,20,21]. Google data are posted publicly, and data can be downloaded directly through Google's web portals or through freely available software. Google normalizes search data by region (in this case, the United States). To normalize the data, each data point is divided by the total searches of the geographic and time range it represents to compare relative popularity; the results are then scaled from a range of 0 to 100 based on a topic's proportion to all searches on all topics [32]. Search terms with low interest appear as "0," while "100" reflects peak interest during a given period. To obtain an unbiased measurement of interest in *coronavirus* and *COVID fatigue*, interest in both search terms was pulled for the duration of the entire study (September 7, 2017, to June 25, 2020). Data received from Google Trends were reported on a weekly scale; interest was reported for each Sunday. To gather a more finely grained real-time view of interest, the data was linearly interpolated using the "approxm" function from the Freqprof package in RStudio.

Data Processing, Modeling, and Visualization in RStudio

Modeling was implemented in lme4 [33] and lmerTest [34]. All plots were generated using ggplot2 [35]. Result tables were produced using stargazer [36]. On average, 164/217 participants (75.6%) reported mobile sensing data each day. For all participants with data, for the following behaviors, inferred through mobile sensing, daily group-level averages were calculated, standardized, and plotted over the time course of January to July 2020: locations visited, phone use, sleep, and time still (Figure 2). The objective of the analyses was to determine how the shock and sustained nature of the pandemic impacted the participants' behaviors and mental health. The term week was modeled as linear and quadratic factors. To obtain a variable mirroring interest in *coronavirus* and *COVID fatigue*, topic interest in both *coronavirus* and *COVID fatigue* was gathered from Google Trends. Interest levels were modeled with fixed effects of reported depression and anxiety levels, unlock duration, unlock number, sedentary time, sleep duration, number of locations visited, and term week (linear and quadratic) variables, as well as random intercepts for each participant. For each variable, any days with missing data (14.47%) were excluded from the analysis. Each variable except participant was scaled to aid model convergence. *P* values were calculated using the Satterthwaite method as implemented in lmerModLmerTest as part of lmerTest.

Figure 2. Standardized daily scores of the number of locations visited, duration of time phone unlocked, sleep duration, and sedentary time before and after the World Health Organization declared COVID-19 a global pandemic (March 11, 2020). All behaviors were calculated using data from the StudentLife app.

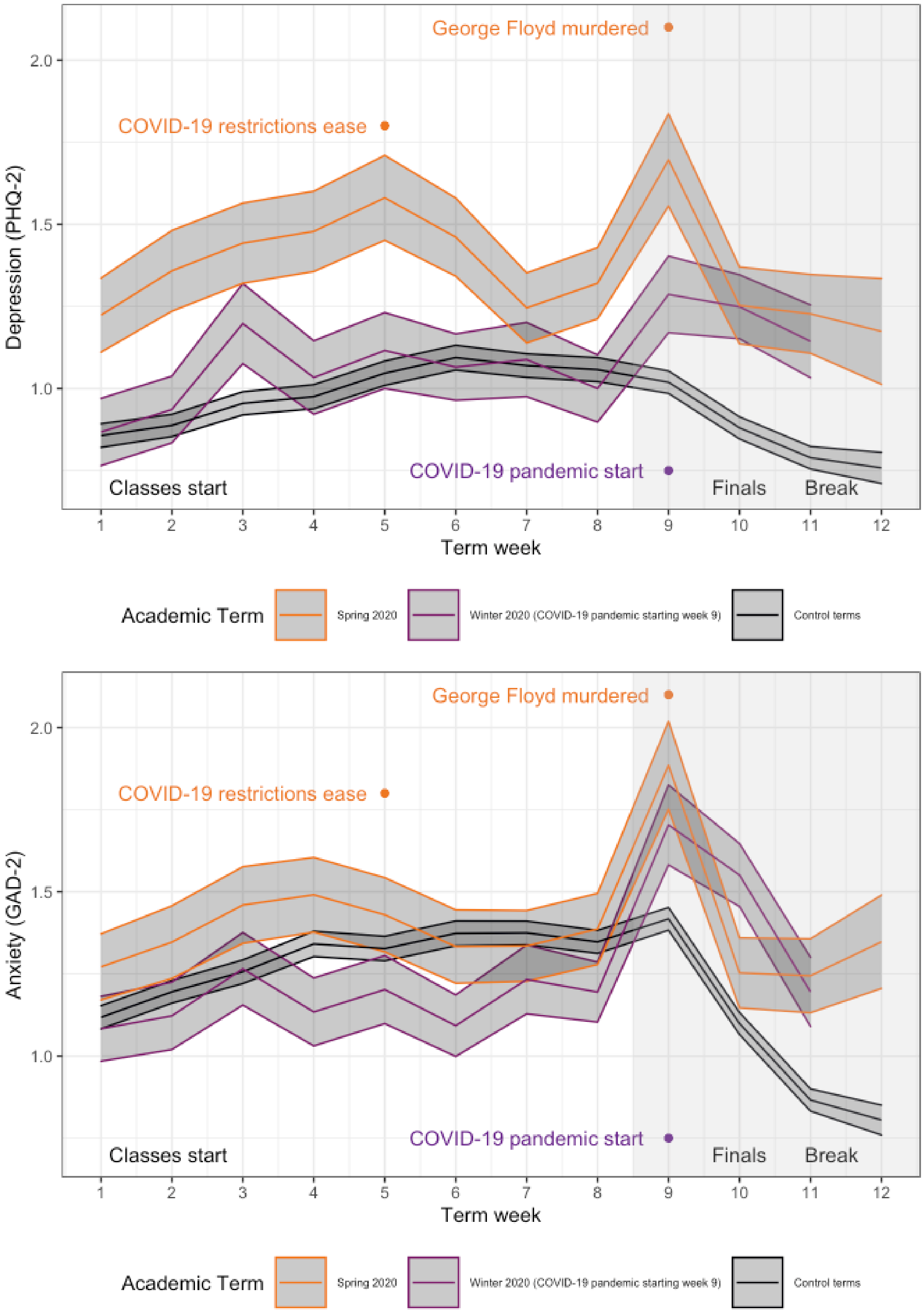


Results

Behavioral and Mental Health Changes Associated With the COVID-19 Pandemic

To visually observe shifts in behaviors due to the COVID-19 pandemic, standardized scores were produced for each behavior plotted with a vertical line marking the date on which the WHO named COVID-19 a pandemic (Figure 2). Students visited fewer locations, used their phones more, spent more time sedentary, and initially exhibited increases in sleep before a sustained decrease in sleep. Notably, the graph shows less extreme peaks and valleys, suggesting that differences in behavior between weekdays and weekends were less pronounced during the COVID-19 pandemic. The mean values of depression and anxiety were plotted by week of the term; data were combined from all study participants from all terms prior to the Winter 2020 term, which was plotted as a separate line along with the Spring 2020 term. Standard error was plotted as a shaded ribbon surrounding the mean. Sustained increased levels of depression and anxiety throughout the Spring 2020 academic term compared to Winter 2020 and control terms were plotted (Figure 3). Peak self-reported anxiety and depression symptoms were observed during the ninth week of the Spring term, corresponding to the time before final examinations, increases in COVID-19 cases, and the murder of George Floyd. Linear models were produced to better assess these differences in a more quantitative manner.

Figure 3. Depression (top) and anxiety (bottom) scores across an academic term and the first 2 weeks of break, with the term influenced by the outbreak of the COVID-19 pandemic as a separate line. The shaded ribbons represent the standard error for each week. Winter 2020 weeks influenced by policy changes related to COVID-19 are represented with a shaded box from weeks 9 to 12. Colored text corresponds to the term in which the event took place. Depression and anxiety were measured with the PHQ-2 and GAD-2 scales through the StudentLife app. Control terms include data from the same group of individuals across previous academic terms. Finals: final examinations; GAD-2: Generalized Anxiety Disorder-2; PHQ-2: Patient Health Questionnaire-2.



Multiple models for sedentary time, phone use, sleep, locations visited, depression, and anxiety were tested (see the *Methods* section for specific details). For all mobile sensing variables except phone use, superior fit was observed with the most complex model that included the COVID-19 academic term, linear and quadratic term week trends, the interaction of the COVID-19 academic term with each of the term week trends, and allowance for random intercepts for each participant's data (random effects). For phone use, superior fit (measured by lowest deviance) was observed with the model including the COVID-19 academic term, linear and quadratic term week trends, and the interaction between the linear term week variable and the COVID-19 academic term. Modeling of the COVID-19 academic term compared with non–COVID-19 academic terms identified significantly increased sedentary time, phone use, sleep, depression, and anxiety, along with a decrease in locations visited (*P*<.001; Table 1). Significant interactions of the COVID-19 term and quadratic term week regressor for sedentary time, locations visited, and anxiety (*P*<.001), as well as sleep and depression (*P*=.05), were observed. Significant interactions between the COVID-19 term and the linear term week regressor for locations visited, phone use, and anxiety (*P*<.001) as well as sedentary time (*P*=.01) and depression (*P*=.05) were also observed.

Table 1. Models of sedentary time, locations visited, sleep duration, phone use, depression, and anxiety by week and by presence of COVID-19 during the academic term.

Variable	Sedentary time (1) (91,541 observations)		Locations visited (2) (85,148 observations)		Sleep duration (3) (91,541 observations)		Phone use (4) (91,541 observations)		Depression (5) (21,284 observations)		Anxiety (6) (21,284 observations)	
	Standardized coefficient (SE)	*P* value	Standardized coefficient (SE)	*P* value	Standardized coefficient (SE)	*P* value	Standardized coefficient (SE)	*P* value	Standardized coefficient (SE)	*P* value	Standardized coefficient (SE)	*P* value
COVID-19 Term	0.469 (0.007)	<.001	–0.851 (0.008)	<.001	0.061 (0.008)	<.001	0.406 (0.007)	<.001	0.262 (0.013)	<.001	0.150 (0.013)	<.001
Term week (linear)	–0.056 (0.003)	<.001	–0.218 (0.003)	<.001	0.041 (0.003)	<.001	0.072 (0.003)	<.001	–0.014 (0.006)	.02	–0.042 (0.006)	<.001
Term week (quadratic)	–0.036 (0.003)	<.001	–0.140 (0.003)	<.001	0.060 (0.003)	<.001	0.013 (0.003)	<.001	–0.074 (0.006)	<.001	–0.108 (0.006)	<.001
COVID-19 term: term week (linear)	–0.021 (0.007)	.003	0.121 (0.008)	<.001	–0.005 (0.007)	.50	–0.032 (0.006)	<.001	0.033 (0.013)	.01	0.086 (0.013)	<.001
COVID-19 term: term week (quadratic)	0.025 (0.007)	.001	0.071 (0.008)	<.001	–0.017 (0.007)	.02	N/A[a]	N/A	0.029 (0.013)	.02	0.088 (0.013)	<.001
Constant	–0.099 (0.033)	.003	0.148 (0.025)	<.001	0.010 (0.032)	.75	–0.102 (0.043)	.02	–0.011 (0.046)	.82	0.031 (0.045)	.49

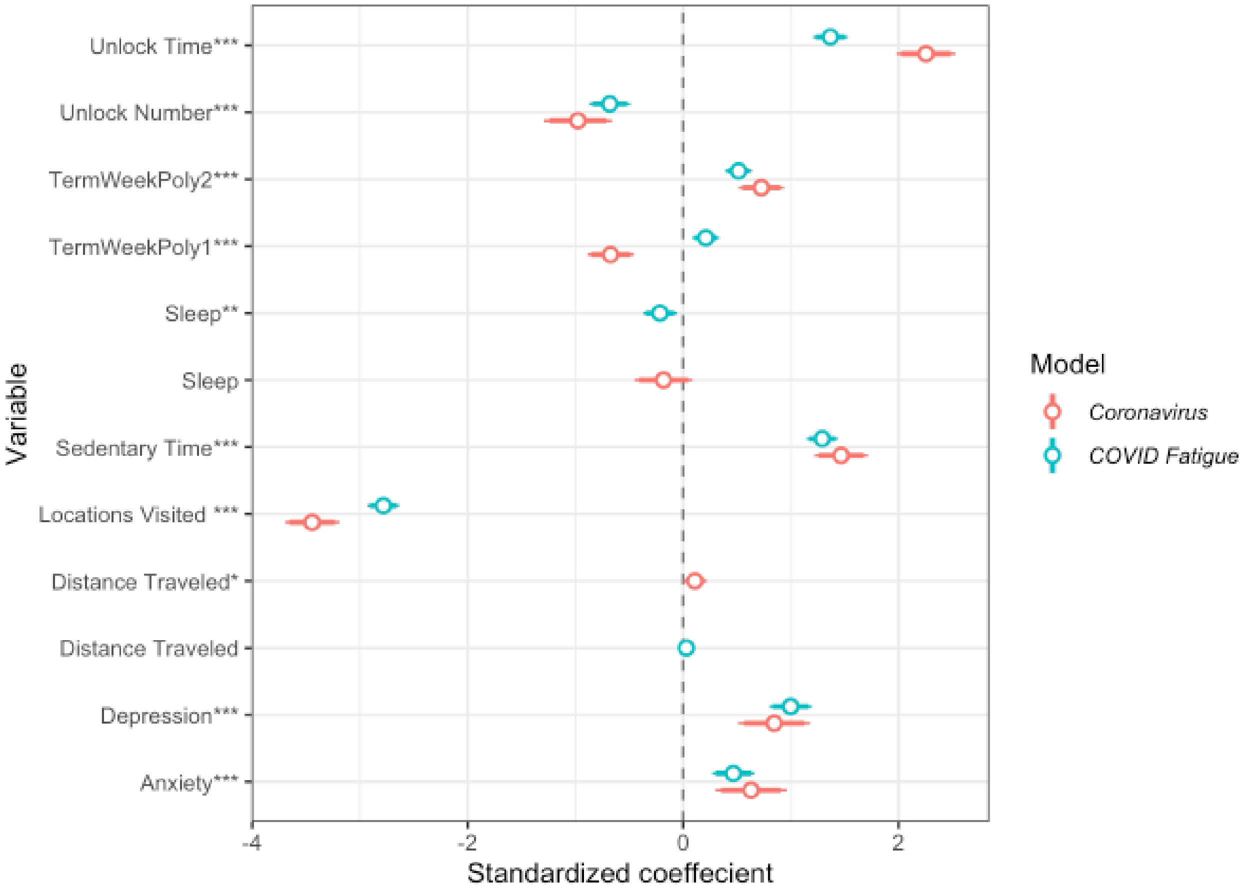
[a] N/A: not applicable.

COVID-19 Interest, Mental Health, and Mobile Sensing

After observing broad differences in sleep, phone use, locations visited, sedentary time, depression, and anxiety between the Winter 2020 term and previous terms, the next goal was to see if these behaviors changed on a daily scale, particularly mirroring the relative interests in *coronavirus* and *COVID fatigue* over the duration of the study. Interest in both variables was at zero until March 2020 (Figure 1). To gather what behaviors changed with *coronavirus* interest and how they changed, we included fixed effects for phone use (unlock duration and unlock number), sedentary time, sleep duration, number of locations visited, and linear and quadratic academic term week regressors. Random intercepts per participant were included in the model. Each variable was scaled to help with convergence of the restricted maximum likelihood model and to obtain regression coefficients that can be compared for relative importance. Models inferring interest in *COVID fatigue* had a better fit (lower deviance) than models of *coronavirus* interest. All variables were significantly associated with interest in the search term *coronavirus*, and all variables except distance traveled were significantly associated with interest in *COVID fatigue*. In both models, phone use (unlock duration) had the largest positive standardized coefficient, followed by sedentary time. The number of locations visited had the largest magnitude of negative standardized coefficient, followed by the number of phone unlocks and sleep duration (*P*<.001).

In the second set of models of interest data, we again made inferences with mobile sensing features, this time with the addition of self-reported anxiety and depression scores. Again, all variables seemed to produce a better fit when inferring interest in COVID fatigue. When anxiety and depression were added to the previously used sensing model, our model demonstrated that increased anxiety and depression were significantly associated with rising interest in *coronavirus* and *COVID fatigue* (*P*<.001). Again, we observed increased phone use (unlock duration), increased sedentary time, and decreased number of locations visited, with standardized beta weights stable across both models. Sleep was negatively associated with interest in *COVID fatigue* (*P*=.05) but no longer had a significant association with *coronavirus* interest, perhaps due to the weekly measured mental health metrics accounting for variance. Notably, both the linear and quadratic term week variables were positively associated with *COVID fatigue* interest, though only the quadratic term week variables were positively associated with *coronavirus* interest. A coefficient plot depicting the relative impacts on behaviors and mental health is shown in Figure 4.

Figure 4. Coefficient plots from mixed linear models of interest in the search terms *coronavirus* and *COVID fatigue*, inferred by mobile smartphone features and self-reported mental health. Intercepts and random intercepts per participant are not plotted. **P*<.05; ***P*<.01; ****P*<.001.



Discussion

Principal Findings

As in our previous work, we found increased interest in COVID-19 to be associated with increased depression and anxiety alongside stark behavioral changes. Further linear models showed that as initial interest in the search term *coronavirus* increased, a cohort of 217 Dartmouth College undergraduate students became more sedentary, visited fewer locations, slept less, used their phones more, and showed increases in anxiety and depression. Although initial models suggested an increase in sleep during academic terms affected by COVID-19, controlling for other sensing variables revealed that students exhibited decreases in sleep as the pandemic endured. Although we observed an increase in distance traveled as *coronavirus* interest rose, we attribute this finding to students returning home, as it coincides with Dartmouth College closing its campus. We found that the search term *coronavirus* had a stronger relationship with changes in behaviors and mental health, with the exception of levels of depression. With sustained adherence to "Stay Home and Stay Safe" guidelines implemented by many national and international governments, we found that as the pandemic endured, students were more sedentary, visited fewer locations (as inferred from GPS tracking), slept less, used their phones more, and exhibited increases in anxiety and depression. Notably, over the entire course of the study, self-reported anxiety and depression levels were highest in the week after George Floyd was murdered. Future digital phenotyping work should investigate the impact of current events on mental health and behavior.

Our models revealed that behaviors and mental health changed in response to the pandemic over a sustained period of time. Although this work is descriptive by nature, behavioral changes sometimes drive changes in mental health in a lagged manner. Notably, a comparison of contemporaneous correlations of data collected prior to and during the COVID-19 pandemic revealed a decrease in the strength of the correlations between sedentary time and mental health (anxiety and depression). Notably, the correlation between locations visited and anxiety also became weaker. Future COVID-19–related work investigating causal relationships between general daily travel, particular sedentary behaviors, and changes in mental health would be of high interest.

Unfortunately, we remain engaged in a battle against the continued spread of COVID-19, necessitating long-term adherence to physical distancing and other mitigation protocols. Although the initial impact of these unprecedented shifts in behavior has been documented, the effects of the longevity of the pandemic are less known. This work uses longitudinal data from a cohort of 217 college students to observe what behavioral and mental health changes are associated with the enduring pandemic. As in previous work, we found increased interest in COVID-19 to be associated with increased depression and behavioral changes. Further linear models showed that as initial interest in *coronavirus* increased, participants became more sedentary, visited fewer locations, slept less, used their phones more, and showed increases in anxiety and depression. Although initial models suggested an increase in sleep during academic terms affected by the COVID-19 pandemic, controlling for other sensing variables revealed that students exhibited decreases in sleep as the pandemic endured.

As in previous work [6], we again observed that COVID-19 is associated with decreased sleep, increased sedentary time, and negative mental health outcomes. Recent work, collected prepandemic, suggests that sleep may mediate a relationship between physical activity and stress [37]. An overwhelming body of research supports the claim that increased physical activity is associated with longer, better quality sleep [38-40]. On the other hand, physical inactivity tends to be associated with increased stress [41], and recent work suggests that exercising during the pandemic may help reduce levels of stress [42]. Stress has been implicated in changes in mental health [43,44] and poor sleep [45-47]. Importantly, other work identifies disturbances in sleep and physical activity that coincide with COVID-19 fatigue [12]. Although the observed prolonged decreases in sleep may potentially mediate a relationship between sustained decreased physical activity and mental well-being during the pandemic, mental health research currently remains uncertain.

Our finding of sustained increased phone use during the COVID-19 pandemic presents a quandary. Work prior to the pandemic suggests that smartphone overuse is associated with negative changes in mental health [48]. However, at a time of widespread social distancing, it seems that smartphones may have a more complex relationship with mental health. Recent research has shown that using smartphones to connect with friends and colleagues may be associated with increased well-being during the pandemic [49]. Moreover, social connection may provide resilience in battling COVID-19 fatigue [50]. Simultaneously, increased phone use during the pandemic may reflect increased social media use and negative changes in mental health [49,51]. Given the observance of increased phone use among college students during the COVID-19 pandemic, future work may seek to uncover the various impacts of phone use during the pandemic.

Primary takeaways from the models are that individuals are maintaining previously seen changes in behavior and mental health. This finding suggests the importance of prioritizing communication with loved ones, and it emphasizes the benefit of maintaining a regular sleep schedule and physical activity to help improve mental health.

Study Limitations and Future Directions

One possible limitation of this study is the search terms used. Although other literature reports use the term “COVID fatigue” to refer to behavioral exhaustion, it is worth noting that similar search terms listed on Google Trends are related to the symptomatology of COVID-19 (ie, fatigue related to virus contraction) rather than to behavioral fatigue. Our graphs show that interest in *COVID fatigue* begins to increase well into the pandemic, but after interest in *coronavirus* increases; however, it is possible that we are partially capturing interest in symptoms. Still, we observed that behaviors and mental health model interest in *covid fatigue* with a better fit than with interest in the term *coronavirus*; this finding suggests that the gradually building interest in COVID-19 fatigue accurately reflects the continuing phases, while interest in the disease itself reflects the initial phase. Future work may consider identifying trending search terms such as *covid mutation* or *covid vaccine* to help quantify the interest in the continuation of the pandemic and how that interest may be impacting mental health. Similarly, as vaccines are more widely distributed, using the terms *covid vaccine* and *herd immunity* along with the terms used in this work could be an effective way to model the impacts of later and (hopefully) final stages of the COVID-19 pandemic on mental health.

Although the mobile smartphone sensing data gathered by the StudentLife app are quite robust, they too have potential limitations. When mobility is decreased, such as during a stay-at-home order, people may not always have their phones with them, which could lead to an overestimation of sedentary time and sleep. Our moderate sample size of 217 total participants across all terms and 212 participants during the COVID-19 pandemic (97.6%) limits the generalizability of the findings. Further, participants may be preferentially accessing larger screens (eg, tablets, laptops, or televisions). Because the app was only installed on the students’ mobile phones, our metric of phone use (as measured by screen unlock duration or number of unlocks) may underestimate the total amount of screen time. However, we did observe increased phone use, suggesting that the effect sizes we observe are underestimated. Future digital phenotyping work capable of distinguishing different types of screen time during the pandemic would be of high interest, as would work that distinguishes the purposes (ie, working, watching Netflix, reading) and proportions of screen use. Further work could also use smartwatches to improve the measurement of behaviors such as sedentary time and allow for more frequent sampling of phone use, location and other measures. Additionally, the use of smartwatches could facilitate highly accurate inferences of time spent exercising, sleep and sedentary time, as these watches typically track heart rate. Despite the moderate sample size and potential data inaccuracies, strong significant effects on mental health and behaviors were observed, suggesting robust effects.

Although finding ways to combat COVID-19 fatigue is beyond the scope of this work, we encourage more research on the subject. In our work, we were unable to infer daily changes in mental health, as to maximize long-term retention, the StudentLife app prompted participants to respond to weekly, not daily, EMAs. Although we presume that college students are likely to be more susceptible to experiencing pandemic fatigue, we do not directly gather fatigue-related feedback. Future studies may consider addressing fatigue more directly, although one should be cautious when choosing metrics of fatigue, as they may also be symptoms of depression (eg, physical exhaustion, anhedonia, trouble sleeping). Administering EMAs on a daily, rather than weekly, basis could help parse day-to-day changes in mental health related to fatigue. In light of the drastic changes in sedentary time, it is critical to find day-to-day activities or routines that may help prevent or reduce fatigue levels. There may be differences in the degree of reported mental health changes between those who spend more time watching television compared to those using social media. Although gathering enough participants may be challenging, using questionnaires to investigate specific behaviors that drive or reduce fatigue levels is critical.

Comparison With Prior Work

In the initial stage of the COVID-19 pandemic, as the media presence of COVID-19 increased, individuals exhibited increased levels of depression and anxiety alongside increased sedentary time, phone use, and a decrease in the number of locations visited. This work examined if the same cohort of individuals exhibited the same changes in behaviors and mental health over the course of a 12-week academic term. Rather than using media presence, this work uses search term interest gathered from Google Trends, a tactic that has recently proved fruitful during the pandemic [9,20]. In this work, we chose the search term *COVID fatigue* to measure the intensity of the pandemic at a given moment to add to the prior knowledge gained during the initial phases of the pandemic. This tactic enabled us to observe what behavioral and mental health changes were associated with the ongoing COVID-19 pandemic, as opposed to focusing on the initial phase, as in prior work. In addition to previously observed shifts in mental health and behaviors, we observed significant decreases in sleep as the COVID-19 pandemic endured.

Conclusions

While the long fight against COVID-19 continues, we must further our understanding of its impacts on mental health. This study provides additional insight into mental health and related behaviors during the ongoing COVID-19 pandemic. Using mixed linear models of smartphone mobile sensing and self-reported mental health questions, we were able to infer the initial and ongoing phases of the COVID-19 pandemic, as measured by search term interest in *coronavirus* and *COVID fatigue*; moreover, we were able to validate that the participants' mental health and related behaviors changed contemporaneously with relative interest levels in these terms. Increases in depression, anxiety, sedentary time, and phone use, alongside decreases in sleep and the number of locations visited, are significantly associated with the ongoing pandemic. All of these changes seem very similar to those seen during the initial phases of the pandemic. As the virus continues to mutate, in future work, we may attempt to identify behaviors as resilience factors for changes in mental health. Similarly, work tracking the distribution of vaccines may elucidate a return to pre–COVID-19 pandemic levels of behavior and mental health.

Acknowledgments

This work was supported by the National Institute of Mental Health (5R01MH059282).

Conflicts of Interest

None declared.

References

1. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
3. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
4. Hasan N, Bao Y. Impact of "e-Learning crack-up" perception on psychological distress among college students during COVID-19 pandemic: a mediating role of "fear of academic year loss". Child Youth Serv Rev 2020 Nov;118:105355 [FREE Full text] [doi: 10.1016/j.childyouth.2020.105355] [Medline: 32834276]
5. Zhang Y, Zhang H, Ma X, Di Q. Mental health problems during the COVID-19 pandemics and the mitigation effects of exercise: a longitudinal study of college students in China. Int J Environ Res Public Health 2020 May 25;17(10):A [FREE Full text] [doi: 10.3390/ijerph17103722] [Medline: 32466163]
6. Huckins J, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal S, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
7. Wang Z, Yang H, Yang Y, Liu D, Li Z, Zhang X, et al. Prevalence of anxiety and depression symptom, and the demands for psychological knowledge and interventions in college students during COVID-19 epidemic: A large cross-sectional study. J Affect Disord 2020 Oct 01;275:188-193 [FREE Full text] [doi: 10.1016/j.jad.2020.06.034] [Medline: 32734907]
8. Liu X, Liu J, Zhong X. Psychological state of college students during COVID-19 epidemic. SSRN Journal. Preprint posted online on March 24, 2020. [doi: 10.2139/ssrn.3552814]
9. Jacobson NC, Lekkas D, Price G, Heinz MV, Song M, O'Malley AJ, et al. Flattening the mental health curve: COVID-19 stay-at-home orders are associated with alterations in mental health search behavior in the United States. JMIR Ment Health 2020 Jun 01;7(6):e19347 [FREE Full text] [doi: 10.2196/19347] [Medline: 32459186]

10. Harvey N. Behavioral fatigue: real phenomenon, naı¨ve construct, or policy contrivance? Front Psychol 2020;11:589892 [FREE Full text] [doi: 10.3389/fpsyg.2020.589892] [Medline: 33224078]
11. Bartoszek A, Walkowiak D, Bartoszek A, Kardas G. Mental well-being (depression, loneliness, insomnia, daily life fatigue) during COVID-19 related home-confinement-a study from Poland. Int J Environ Res Public Health 2020 Oct 12;17(20):7417 [FREE Full text] [doi: 10.3390/ijerph17207417] [Medline: 33053787]
12. Field T, Mines S, Poling S, Diego M, Bendell D, Veazey C. COVID-19 lockdown fatigue. American Journal of Psychiatric Research and Reviews 2021;4(1):27. [doi: 10.28933/ajprr-2020-12-0906]
13. Rahmandad H, Lim TY, Sterman J. Behavioral dynamics of COVID-19: estimating underreporting, multiple waves, and adherence fatigue across 92 nations. Syst Dyn Rev 2021 Mar 16;37(1):5-31. [doi: 10.1002/sdr.1673]
14. Morgul E, Bener A, Atak M, Akyel S, Aktaş S, Bhugra D, et al. COVID-19 pandemic and psychological fatigue in Turkey. Int J Soc Psychiatry 2020 Jul 10:20764020941889 [FREE Full text] [doi: 10.1177/0020764020941889] [Medline: 32650681]
15. Carlucci L, D'Ambrosio I, Balsamo M. Demographic and attitudinal factors of adherence to quarantine guidelines during COVID-19: the Italian model. Front Psychol 2020;11:559288 [FREE Full text] [doi: 10.3389/fpsyg.2020.559288] [Medline: 33192820]
16. Ye B, Zhou X, Im H, Liu M, Wang X, Yang Q. Epidemic rumination and resilience on college students' depressive symptoms during the COVID-19 pandemic: the mediating role of fatigue. Front Public Health 2020;8:560983 [FREE Full text] [doi: 10.3389/fpubh.2020.560983] [Medline: 33363075]
17. Faurholt-Jepsen M, Busk J, Þórarinsdóttir H, Frost M, Bardram J, Vinberg M, et al. Objective smartphone data as a potential diagnostic marker of bipolar disorder. Aust N Z J Psychiatry 2019 Feb;53(2):119-128. [doi: 10.1177/0004867418808900] [Medline: 30387368]
18. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun 28;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
19. Insel T. Digital phenotyping: technology for a new science of behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
20. Hoerger M, Alonzi S, Perry L, Voss H, Easwar S, Gerhart J. Impact of the COVID-19 pandemic on mental health: real-time surveillance using Google Trends. Psychol Trauma 2020 Sep;12(6):567-568. [doi: 10.1037/tra0000872] [Medline: 32790441]
21. Tana J, 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]
22. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
23. Fried EI, Papanikolaou F, Epskamp S. Mental health and social contact during the COVID-19 pandemic: an ecological momentary assessment study. PsyArXiv. Preprint posted online on April 24, 2020. [doi: 10.31234/osf.io/36xkp]
24. StudentLife study. Dartmouth College. URL: https://studentlife.cs.dartmouth.edu/ [accessed 2021-05-26]
25. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: UbiComp '14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Sep Presented at: 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 2014; Seattle, WA. [doi: 10.1145/2632048.2632054]
26. Huckins JF, daSilva AW, Wang R, Wang W, Hedlund EL, Murphy EI, et al. Fusing mobile phone sensing and brain imaging to assess depression in college students. Front Neurosci 2019 Mar 21;13:248 [FREE Full text] [doi: 10.3389/fnins.2019.00248] [Medline: 30949024]
27. Kroenke K, Spitzer R, Williams J, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ–4. Psychosomatics 2009 Nov;50(6):613-621. [doi: 10.1016/s0033-3182(09)70864-3]
28. ActivityRecognitionApi. Google Play services. URL: https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionApi [accessed 2021-05-26]
29. Wang R, Aung M, Abdullah S, Brian R, Campbell A, Choudhury T, et al. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In: UbiComp '16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Sep Presented at: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 12-16, 2016; Heidelberg, Germany p. 886-897. [doi: 10.1145/2971648.2971740]
30. IOS core motion. Apple Developer. URL: https://developer.apple.com/reference/coremotion [accessed 2021-05-26]
31. Daszykowski M, Walczak B. Density-based clustering methods. In: Comprehensive Chemometrics (Second Edition). Amsterdam, Netherlands: Elsevier; 2009:565-580.
32. Google Trends. URL: https://www.google.com/trends [accessed 2021-05-26]
33. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models Using lme4. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
34. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Soft 2017;82(13). [doi: 10.18637/jss.v082.i13]
35. Wickham H. ggplot2. WIREs Comp Stat 2011 Feb 04;3(2):180-185. [doi: 10.1002/wics.147]

36. Hlavac M. Stargazer: well-formatted regression and summary statistics tables. R Project. 2018. URL: https://cran.r-project.org/web/packages/stargazer/index.html [accessed 2021-05-26]
37. Zhai X, Wu N, Koriyama S, Wang C, Shi M, Huang T, et al. Mediating effect of perceived stress on the association between physical activity and sleep quality among Chinese college students. Int J Environ Res Public Health 2021 Jan 02;18(1):289 [FREE Full text] [doi: 10.3390/ijerph18010289] [Medline: 33401720]
38. Kredlow M, Capozzoli M, Hearon B, Calkins A, Otto M. The effects of physical activity on sleep: a meta-analytic review. J Behav Med 2015 Jun;38(3):427-449. [doi: 10.1007/s10865-015-9617-6] [Medline: 25596964]
39. Lang C, Brand S, Feldmeth AK, Holsboer-Trachsler E, Pühse U, Gerber M. Increased self-reported and objectively assessed physical activity predict sleep quality among adolescents. Physiol Behav 2013 Aug 15;120:46-53. [doi: 10.1016/j.physbeh.2013.07.001] [Medline: 23851332]
40. Loprinzi PD, Cardinal BJ. Association between objectively-measured physical activity and sleep, NHANES 2005–2006. Ment Health Phys Act 2011 Dec;4(2):65-69. [doi: 10.1016/j.mhpa.2011.08.001]
41. Zhai X, Ye M, Wang C, Gu Q, Huang T, Wang K, et al. Associations among physical activity and smartphone use with perceived stress and sleep quality of Chinese college students. Ment Health Phys Act 2020 Mar;18:100323. [doi: 10.1016/j.mhpa.2020.100323]
42. Deng C, Wang J, Zhu L, Liu H, Guo Y, Peng X, et al. Association of web-based physical education with mental health of college students in Wuhan during the COVID-19 outbreak: cross-sectional survey study. J Med Internet Res 2020 Oct 05;22(10):e21301 [FREE Full text] [doi: 10.2196/21301] [Medline: 32997639]
43. van Praag H. Can stress cause depression? Prog Neuropsychopharmacol Biol Psychiatry 2004 Aug;28(5):891-907. [doi: 10.1016/j.pnpbp.2004.05.031] [Medline: 15363612]
44. Huckins J, DaSilva A, Hedlund E, Murphy E, Rogers C, Wang W, et al. Causal factors of anxiety and depression in college students: longitudinal ecological momentary assessment and causal analysis using Peter and Clark momentary conditional independence. JMIR Ment Health 2020 Jun 10;7(6):e16684 [FREE Full text] [doi: 10.2196/16684] [Medline: 32519971]
45. Caldwell K, Harrison M, Adams M, Quin RH, Greeson J. Developing mindfulness in college students through movement-based courses: effects on self-regulatory self-efficacy, mood, stress, and sleep quality. J Am Coll Health 2010 Oct 22;58(5):433-442 [FREE Full text] [doi: 10.1080/07448480903540481] [Medline: 20304755]
46. Lee S, Wuertz C, Rogers R, Chen Y. Stress and sleep disturbances in female college students. Am J Hlth Behav 2013 Nov 01;37(6):851-858. [doi: 10.5993/ajhb.37.6.14]
47. Amaral A, Soares M, Pinto A, Pereira A, Madeira N, Bos S, et al. Sleep difficulties in college students: the role of stress, affect and cognitive processes. Psychiatry Res 2018 Feb;260:331-337. [doi: 10.1016/j.psychres.2017.11.072] [Medline: 29227897]
48. Demirci K, Akgönül M, Akpinar A. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. J Behav Addict 2015 Jun;4(2):85-92 [FREE Full text] [doi: 10.1556/2006.4.2015.010] [Medline: 26132913]
49. Sun R, Rieble C, Liu Y, Sauter D. Connected despite lockdown: the role of social interactions and social media use in wellbeing. PsyArXiv. Preprint posted online on December 30, 2020. [doi: 10.31234/osf.io/x5k8u]
50. Nitschke J, Forbes P, Ali N, Cutler J, Apps M, Lockwood P, et al. Resilience during uncertainty? Greater social connectedness during COVID-19 lockdown is associated with reduced distress and fatigue. Br J Health Psychol 2021 May;26(2):553-569. [doi: 10.1111/bjhp.12485] [Medline: 33099800]
51. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020 Apr 16;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]

Abbreviations

DBSCAN: density-based spatial clustering of applications with noise
EMA: ecological momentary assessment
GAD-2: Generalized Anxiety Disorder-2
PHQ-2: Patient Health Questionnaire-2
PHQ-4: Patient Health Questionnaire-4
WHO: World Health Organization

Edited by C Basch; submitted 17.03.21; peer-reviewed by Y Zhang, A Teles; comments to author 04.04.21; revised version received 21.04.21; accepted 22.04.21; published 04.06.21.

Please cite as:
Mack DL, DaSilva AW, Rogers C, Hedlund E, Murphy EI, Vojdanovski V, Plomp J, Wang W, Nepal SK, Holtzheimer PE, Wagner DD, Jacobson NC, Meyer ML, Campbell AT, Huckins JF
Mental Health and Behavior of College Students During the COVID-19 Pandemic: Longitudinal Mobile Smartphone and Ecological Momentary Assessment Study, Part II
J Med Internet Res 2021;23(6):e28892
URL: https://www.jmir.org/2021/6/e28892
doi:10.2196/28892
PMID:33900935

Original Paper

Coping Styles for Mediating the Effect of Resilience on Depression Among Medical Students in Web-Based Classes During the COVID-19 Pandemic: Cross-sectional Questionnaire Study

Lina Zhao[1], MA; Kristin Sznajder[2], PhD; Dan Cheng[3], MA; Shimeng Wang[4], BA; Can Cui[5], BA; Xiaoshi Yang[5], PhD

[1]Institute of Foreign Languages, China Medical University, Shenyang, China
[2]Department of Public Health Sciences, College of Medicine, Pennsylvania State University, Philadelphia, PA, United States
[3]Hospital of Stomatology, China Medical University, Shenyang, China
[4]Department of Anesthesiology, Shengjing Hospital of China Medical University, Shenyang, China
[5]Department of Social Medicine, College of Health Management, China Medical University, Shenyang, Liaoning Province, China

Corresponding Author:
Xiaoshi Yang, PhD
Department of Social Medicine
College of Health Management
China Medical University
No 77 Puhe Road, Shenyang North New Area
Shenyang, Liaoning Province, 110122
China
Phone: 86 18900910796
Email: xsyang@cmu.edu.cn

Abstract

Background: Due to strict, nationwide, comprehensive COVID-19 protective measures, including home quarantine, all Chinese medical students began taking web-based classes beginning in the spring semester of 2020. Home quarantine, web-based classes, and the stress surrounding the COVID-19 pandemic may have triggered an increased incidence of mental health problems among medical students. Although there have been increasing amounts of literature on depression among medical students, studies focusing on positive psychological resources, such as resilience during the COVID-19 pandemic, still need to be expanded.

Objective: This study aims to assess depression among medical students who are taking web-based classes during the COVID-19 pandemic and to investigate the role of coping styles as mediators between resilience and depression.

Methods: A cross-sectional study of 666 medical students involving stratified sampling in Shenyang, Liaoning Province, China, was completed between March 20 and April 10, 2020. The participants responded to a self-administered, smartphone-based questionnaire, which included the Patient Health Questionnaire-9, Simplified Coping Style Questionnaire, and Ego Resilience 89 Scale. Hierarchical linear regression and structural equation modeling were used in this study.

Results: The prevalence of depression among the participants was 9.6% (64/666) in this study. The regression analysis revealed that grade (the year in which the medical student was in training) (P=.013), how well students adapted to web-based classes (P<.001), their levels of resilience (P=.04), and their coping styles were independent predictors for depression (P<.001). Resilience and positive coping styles were negatively related to depression (resilience: P=.04; positive coping styles: P<.001), and negative coping styles were positively related to depression (P<.001). The structural equation modeling analysis showed that the effect of resilience on depression was partially mediated by coping styles (P=.007).

Conclusions: In this study, it was found that the prevalence of depression was slightly low and coping styles mediated the association between resilience and depression among medical students during the COVID-19 pandemic. These findings have significant implications for future studies. Future studies and interventions should aim to improve resilience and promote positive coping styles.

(J Med Internet Res 2021;23(6):e25259) doi:10.2196/25259

KEYWORDS

resilience; coping styles; depression; medical students; COVID-19; coping; mediation; web-based education; e-learning; smartphone; cross-sectional

Introduction

The COVID-19 pandemic has deeply affected peoples’ lives all over the world since its emergence in 2019. Both isolation and economic pressure have had a profound impact on the psychosocial environment in each affected country. The pandemic has also increased the public’s susceptibility to detrimental psychological consequences [1]. According to a Chinese study, more than half the study population reported moderate or severe psychological impacts resulting from the COVID-19 pandemic [2]. Quarantine-associated mental health issues involve depression, anxiety, and irritability [3].

School lockdowns were implemented in many countries [4]. In order to better prevent and control the further spread of COVID-19, many universities in China began conducting web-based classes in February 2020. With the unprecedented number of web-based classes, home-quarantined medical students' psychological status is an important area of investigation. It has been reported that the rate of anxiety and depression among medical students is high [5], and the COVID-19 pandemic might bring about further risks to their mental health. Medical courses usually involve practice and experiments. However, web-based classes cannot provide such opportunities, which may add to medical students' worries about their academic achievements and result in high amounts of mental stress. Students have been reported to prefer studying in classrooms over having web-based classes due to the feeling of being together during classroom learning and the ability to share perspectives [6]. Therefore, the lack of peer contact and face-to-face communication with teachers in web-based classes may increase the risk of developing worry, anxiety, or even depression among medical students.

Depression—a mood disorder defined by sadness, inactivity, the loss of appetite or overeating, and difficulty in concentrating—can result in the reduced ability to perform daily activities among some people. Depression is one of the most often identified health issues among college undergraduates [7,8] and is especially common among medical undergraduates [9,10]. Previous studies have found that approximately 30% of medical undergraduates in Europe experience anxiety or depression [11,12]. Mental health disorders among home-quarantined university students have also been found in a previous study, which reported that the prevalence of depression was as high as 9% among university students about 1 month after COVID-19 outbreak in China [13]. A large cross-sectional study that included 44,447 Chinese college students reported an overall prevalence of depression symptoms of 12.2% during the COVID-19 pandemic [14]. The increased prevalence of depression was also observed in the winter 2020 academic semester in a US study [15].

The cognitive-behavioral model of health anxiety [16] suggests that some individuals have maladaptive assumptions about their health and, consequently, tend to overconsume health information, which could lead to high levels of anxiety [17]. Based on this model, individuals who have high levels of anxiety tend to be more anxious during a pandemic [18,19]. The transactional stress model theory [20] states that the responses of individuals who are faced with stress are affected by the coping process during a stressful experience [21] and that situation appraisal and coping can be influenced by positive personal resources.

There are many recent, positive psychology [22,23] studies that focus on depression, and resilience—a positive capability—is a topic of wide concern [24,25]. Resilience refers to one’s capability to adjust to challenges and adverse events [26,27] such as trauma, threats, or other major stresses, and resilience may prevent depression [28-30]. People who are less resilient are more susceptible to pathological reactions to adversities, while people who are more resilient are more likely to be protected against adversities [31].

According to the transactional stress model theory, the process of coping plays a significant role in individuals’ responses to stress [21]. Coping presents strategies of cognition and behavior that individuals can use to master, decrease, or stand up to the inward or outward demands of stressful situations [32]. Dynamic reactions to adversities aid individuals in preventing themselves from developing psychological impairments. Coping styles generally consist of two categories—positive coping and negative coping. Positive coping involves managing problems, adjusting quickly to stressors, and allaying pressure, whereas negative coping includes avoidance, social withdrawal, and the pitying of oneself, which all exacerbate anxiety. Studies have shown the significant correlation between negative coping styles and depression [33,34]. Positive coping styles help individuals to cope with adversities actively. This may involve seeking others’ advice and finding out solutions to problems, which is beneficial for mental well-being. Previous studies have indicated that coping styles play a mediating role in the relationship between perfectionism and depression among undergraduate students [35] and mediate the association between depression and eating disorders among Chinese female undergraduates [36]. Thus, it has been speculated that coping styles would mediate the relationship between resilience and depression.

This study assessed the prevalence of depression among medical students and determined whether coping styles play a mediating role in the relationship between resilience and depression among medical students during the COVID-19 pandemic. This study examined the following three hypotheses: (1) higher levels of resilience predict lower depression scores; (2) coping exerts a positive effect on relieving depression; and (3) coping styles mediate the relationship between resilience and depression.

Methods

Study Design and Participants

A cross-sectional study involving stratified sampling was carried out by conducting a self-administered questionnaire on Wenjuanxing—a smartphone- and web-based questionnaire platform—between March 20 and April 10, 2020. Wenjuanxing is a widely used, open, web-based questionnaire platform that was developed by Changsha Ranxing Information and Technology Limited Company. The free and self-design version was used in this study. The validity and reliability of the questionnaire that we designed and used in this study were examined.

Medical students who were home-quarantined in their first, second, or third year at China Medical University were eligible for this study. In total, 8 classes from each grade in which the medical students were in training were randomly selected. The medical students from 24 classes in their first, second, or third year at China Medical University were selected as the participants and finished the questionnaire. Of the total 720 medical students who were recruited in this study, 666 participants responded completely to the questionnaire, resulting in a valid response rate of 92.5% (666/720).

Ethics Statement

All participants were fully informed of the study protocol and provided informed consent prior to taking the web-based questionnaire. Participation was voluntary and anonymous. The study protocol was approved by the Ethics Committee of China Medical University.

Demographic Characteristics of Participants

Demographic information, including grade (the year in which the medical student was in training; ie, freshman, sophomore, and junior year), gender, age (<20 years and ≥20 years), fathers' education (junior middle school and below and specialized secondary school and above), mothers' education (junior middle school and below and specialized secondary school and above), monthly income (≤RMB 5000 [US $778.30] and >RMB 5000 [US $778.30]), major (clinical medicine and others), and whether students adapted to web-based classes (yes or no), was collected.

Measurement of Depression

Depression was measured with the Patient Health Questionnaire-9 (PHQ-9), which is commonly used for the measurement of depression based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria, is comparably sensitive and specific, and includes 9 items [37]. A score of ≥10 on the PHQ-9 was considered as the indicator for the existence of depression. The PHQ-9 has been widely used in previous studies among the Chinese population [38]. Many previous studies have confirmed that the PHQ-9 has good reliability, ranging from 0.749 to 0.92 [38-41], and the Cronbach α coefficient of the PHQ-9 was .927 in this study.

Measurement of Coping Styles

Coping styles were assessed with the Simplified Coping Style Questionnaire, which included 20 items using a Likert scale of 0 (never) to 3 (frequently). In this study, the Cronbach α coefficient of the Simplified Coping Style Questionnaire was .862.

Measurement of Resilience

Resilience was measured with the Ego Resilience 89 Scale, which has good internal reliability [42] as well as superior construct validity [43]. Participants completed the 14-item, 4-point scale by indicating the degree to which they approved of each statement, with scores ranging from 1 to 4 (1="does not apply at all"; 2="applies slightly, if at all"; 3="applies somewhat"; 4="applies quite strongly"). In this study, the Cronbach α coefficient of the Ego Resilience 89 Scale was .935.

Statistical Analysis

SPSS 17.0 (IBM Corporation) and AMOS (Analysis of Moment Structures) 24.0 (IBM Corporation) were used for statistical analyses in the present study. The comparison of differences among classified groups was conducted with *t* tests (two-tailed) and one-way analysis of variance tests, and a two-tailed *P* value of less than .05 considered statistically significant.

In order to examine the incremental variance of any given set of independent variables and to assess the mediating role of coping styles in the association between resilience and depression, hierarchical linear regression analysis was used. Depression was used as the dependent variable. Resilience and coping styles were used as the independent variables. The variables were entered into models via the following step-by-step process: in step 1, the demographic characteristics of the medical students were entered; in step 2, resilience was entered; and in step 3, coping styles were entered. The following criteria for establishing the mediating effects, according to Baron and Kenny [44] in their approach to analyzing mediation, are supposed to be met: (1) the independent variable (resilience) is significantly related to both the dependent variable (depression) and the mediator (coping styles); (2) the mediator (coping styles) is significantly related to the dependent variable (depression); and (3) the adding of the mediator (coping styles) in the model significantly lessens or clears away the independent variable's (resilience's) impact on the dependent variable (depression).

To prove that coping styles played a mediating role in the relationship between resilience and depression, structural equation modeling was used. Bootstrapping strategies were used to examine the mediating role (a×b product) of coping styles in the relationship between resilience and depression. The bootstrap estimate was based on 5000 bootstrap samples, and the bias-corrected and accelerated (BCa) 95% CI for each a×b product was examined. The goodness of fit was determined by the following: a chi-square to df ratio of <5, a goodness-of-fit index (GFI) of >0.90, a comparative fit index (CFI) of >0.90, a root mean square error of approximation (RMSEA) of <0.08, and a Tucker-Lewis Index (TLI) of >0.90.

Ethical Approval

The study protocol conformed to the ethical standards of and was approved by the Ethic Committee of China Medical University. All participants gave their consent after being informed of the purpose and procedure of the study via a web-based platform. The confidentiality and anonymity of all participants' collected information were ensured.

Informed Consent

The informed consent of every participant was acquired before the launch of the procedures of this research.

Results

Demographic Characteristics and Depression Distribution Among the Participants

The demographic characteristics and their respective mean depression scores are shown in Table 1. Approximately 41.7% (278/666) of study participants were freshmen. The average age of the participants was 20 years. Approximately 39.3% (262/666) of the participants were males. The students reported their fathers' education and mothers' education as junior middle school or below, accounting for 55.4% (369/666) and 59.9% (399/666) of the responses, respectively. About 49.2% (328/666) of the participants reported their family's monthly income as more than RMB 5000 (US $778.30). Most students were studying clinical medicine (460/666, 69.1%). With respect to gender, the depression scores of the male students were significantly higher than those of the female students (*P*=.045). Up to 88.9% (592/666) of the participants were adapting to web-based classes, while 11.1% (74/666) were not. The depression scores among the students who were not adapting to web-based classes were significantly higher than those among the students who were adapting to web-based classes (*P*<.001).

Table 1. Demographic characteristics and the distributions of depression among students (N=666).

Variables	Value, n (%)	Depression score, mean (SD)
Grade		
Freshman	278 (41.7)	4.31 (4.83)
Sophomore and junior	388 (58.3)	3.67 (4.20)
Gender		
Male	262 (39.3)	4.37 (5.12)[a]
Female	404 (60.7)	3.66 (3.98)
Age (years)		
<20	315 (47.3)	3.84 (4.48)
≥20	351 (52.7)	4.02 (4.48)
Fathers' education		
Junior middle school and below	369 (55.4)	4.02 (4.69)
Specialized secondary school and above	297 (44.6)	3.84 (4.21)
Mothers' education		
Junior middle school and below	399 (59.9)	3.99 (4.47)
Specialized secondary school and above	267 (40.1)	3.86 (4.50)
Monthly income (RMB [US $])		
≤5000 (US $778.30)	338 (50.8)	3.84 (4.16)
>5000 (US $778.30)	328 (49.2)	4.04 (4.79)
Major		
Clinical medicine	460 (69.1)	3.91 (4.54)
Others	206 (30.9)	4.00 (4.35)
Adapting to web-based classes		
Yes	592 (88.9)	3.44 (4.07)
No	74 (11.1)	7.88 (5.54)[b]

[a] Significant at the .05 level (two-tailed).

[b] Significant at the .01 level (two-tailed).

Correlations Between Depression and Continuous Variables

The correlations between depression and the continuous variables are shown in Table 2. Depression among medical students was significantly and negatively associated with both resilience (*P*<.001) and positive coping styles (*P*<.001), while depression was significantly and positively associated with negative coping styles among medical students (*P*<.001).

Table 2. The correlations among depression and continuous variables.

Variables[a]	Depression	Resilience	Positive coping styles	Negative coping styles
Depression				
r	1	−0.288[b]	−0.332[b]	0.356[b]
P value	—[c]	<.001	<.001	<.001
Resilience				
r	−0.288[b]	1	0.558[b]	−0.089[b]
P value	<.001	—	<.001	<.001
Positive coping styles				
r	−0.332[b]	0.558[b]	1	0.078[b]
P value	<.001	<.001	—	<.001
Negative coping styles				
r	0.356[b]	−0.089[b]	0.078[b]	1
P value	<.001	<.001	<.001	—

[a] The mean scores for depression, resilience, positive coping styles, and negative coping styles are 3.94 (SD 4.48), 43.88 (SD 7.77), 38.80 (SD 6.76), and 17.79 (SD 4.83), respectively.

[b] Significant at the .01 level (two-tailed).

[c] Not applicable.

The linear regression models of depression among medical students are presented in Table 3. The final regression model (model 3) explained 31% of the total variance in depression scores. Resilience and coping styles explained 7% and 13.5% of the total variance in depression scores, respectively. Grade (*P*=.013), resilience (*P*=.04), positive coping styles (*P*<.001), and negative coping styles (*P*<.001) were significant predictors for depression. Grade, resilience, and positive coping styles were negatively associated with depression, while negative coping styles were positively associated with depression.

Table 3. The hierarchical linear regression analysis of depression.

Variables	Depression, standardized β		
	Model 1[a]	Model 2[b]	Model 3[c]
Block 1: demographic characteristics			
Grade (freshman vs sophomore and junior)	−.088	−.112[d]	−.103[d]
Gender (male vs female)	−.029	−.040	.000
Age (<20 years vs ≥20 years）	.051	.073	.079
Fathers’ education (junior middle school and below vs specialized secondary school and above)	−.019	.018	.022
Mothers’ education (junior middle school and below vs specialized secondary school and above)	−.019	−.022	−.019
Monthly income (≤RMB 5000 [US $778.30] vs >RMB 5000 [US $778.30])	.043	.049	.041
Major (clinical medicine vs others)	−.019	−.037	−.053
Adapting to web-based classes (yes vs no)	.303	.270[d]	.197[e]
Block 2: resilience	—[f]	−.270[e]	−.087[d]
Block 3: coping styles			
Positive coping style	—	—	−.287[e]
Negative coping style	—	—	.335[e]

[a] The R^2 and ΔR^2 values of model 1 are 0.105 and 0.105, respectively.

[b] The R^2 and ΔR^2 values of model 2 are 0.175 and 0.070, respectively.

[c] The R^2 and ΔR^2 values of model are 0.310 and 0.135, respectively.

[d] Significant at the .05 level (two-tailed).

[e] Significant at the .01 level (two-tailed).

[f] Not applicable.

The Mediating Role of Coping Styles in the Relationship Between Resilience and Depression

Figure 1 presents the direct effect of resilience on depression (c=−.34; *P*<.001) before coping styles were entered as a mediator. The model revealed that resilience had a significant negative effect on depression (*P*<.001), and this model had good model fit indices (chi-square to df ratio<5; RMSEA=0.052; CFI=0.962; GFI=0.928; adjusted GFI=0.904; TLI=0.954; Figure 1).

Figure 1. Standardized solutions for the structural equation model of resilience and depression. The standardized path coefficient is shown on the unidirectional arrow path. *The coefficient of the path is significant at the *P*<.05 level.

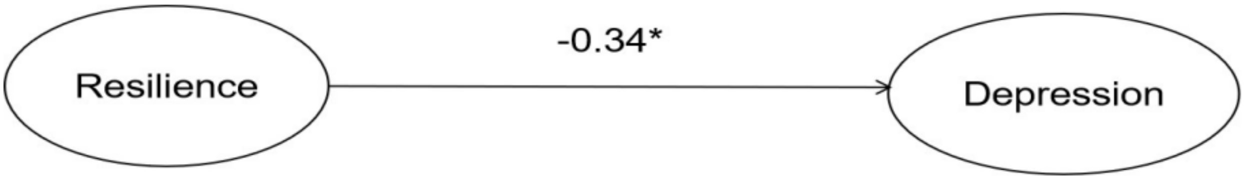
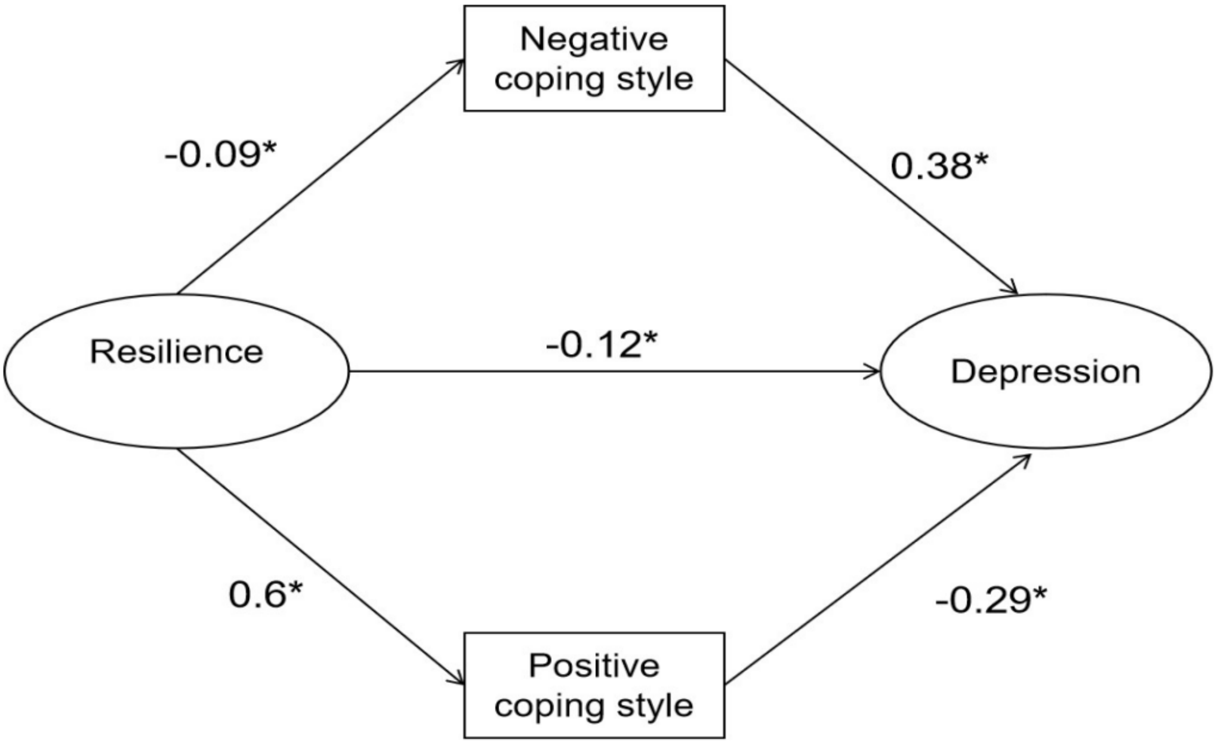


Figure 2 represents the structural equation modeling of the mediating role of coping styles in the relationship between resilience and depression, and the standardized path coefficients are presented on the unidirectional arrow paths. When coping styles were used as the mediator, the path coefficient between resilience and depression decreased significantly (from −.34 in Figure 1 to −.12 in Figure 2; *P*=.007), which confirmed coping styles' partial mediating role in the association between resilience and depression. This model yielded a good model fit (chi-square to df ratio<5; RMSEA=0.051; CFI=0.957; GFI=0.923; adjusted GFI=0.900; TLI=0.949). According to the BCa bootstrap test, coping styles played a significant mediating role in the association between resilience and depression (*P*=.007; a×b=−0.22; BCa 95% CI −0.324 to −0.153), which proved that coping styles had a significant role as a mediator between resilience and depression.

Figure 2. Structural equation modeling of the mediating role of coping styles in the relationship between resilience and depression. Standardized path coefficients are shown on the unidirectional arrow paths. *The coefficient of the path is significant at the *P*<.05 level.



Discussion

As far as we know, this study presents the first attempt to investigate the relationship among resilience, coping styles, and depression among Chinese medical students in the context of web-based classes during the COVID-19 pandemic. In this study, 64 out of the 666 (9.6%) participants reported depression, which was slightly higher than the depression prevalence of 9% in a study of home-quarantined Chinese college students that was conducted during the COVID-19 pandemic [13] but lower than the prevalence rates in most previous studies. Such studies have indicated high prevalence rates of depression among college students (ranging from 12.2% to 25.3%) during the COVID-19 pandemic [14,45,46]. Depression is one of the most commonly occurring mental health issues among college students [47-49], and the prevalence of depression is especially high among medical students [10,50-52]. However, the prevalence of depression among medical students in this study was slightly lower than that of a Swedish study (12.9%) [53] and much lower than that of a study (25%) on US and Canadian medical students [10,54]. One possible reason why this study found lower rates of depression among Chinese medical students may be that staying with family helped ease symptoms of depression. This is good for mental health, and web-based classes provided the students with more opportunities to use available resources at home for entertainment, which might have helped with easing symptoms of depression. Furthermore, due to their medical knowledge, medical students may be more likely to perceive the pandemic objectively, which might prevent depression.

In this study, grade (*P*=.013) and whether students adapted to web-based classes (*P*<.001) had significant impacts on depression. With the national requirement of “suspending classes without suspension of learning” in China during the COVID-19 pandemic, web-based studying at home provided the only feasible method for keeping up with the learning schedule. The results in this study showed that the 11.1% (74/666) of participants who did not adapt to web-based classes had significantly higher depression scores (*P*<.001). This may be because the students who failed to adapt to web-based classes were more accustomed to face-to-face instruction and web-based classes added to their stress, which could result in increased levels of depression.

It was found that resilience was negatively related to depression in this study. With regard to our hypotheses, which were based on the transactional stress model theory [20], the findings from this study indicated that higher levels of resilience were predictive of lower levels of depression. This is in line with a substantial number of previous studies [55-58]. A prospective, multi-institutional study focusing on US medical students reported that resilient students were less susceptible to depression [59]. It was clarified in previous studies that resilience played a role in attenuating depression in different populations [56,57], such as college students and medical and nursing students [60,61]. Our study indicated that resilience played a protective role against depression, which is in line with prior studies [62,63]. Higher levels of resilience have been found to correlate with better subjective health [60], less distress [64], and more optimism among medical students [65]. Previous studies have also found that resilience among medical students might play a critical role in maintaining mental well-being during the COVID-19 pandemic. Web-based classes have been viewed as an intervention that plays a role in decreasing levels of anxiety associated with the pandemic [66]. However, there have been some students who might have had low resilience and failed to adapt to web-based classes well. Lower resilience has been found to be related to higher incidence rates for psychological issues [67,68]. Home-quarantined medical students who transitioned from learning in classrooms on campus to web-based learning at home might have experienced added stress, which might trigger depression. Resilience might help medical students adapt to uncertainty and maintain mental well-being while taking web-based classes at home during the COVID-19 pandemic. First, high resilience might help medical students combat the stressful situations of web-based classes, thus relieving the symptoms of depression. However, students with low resilience might be less likely to adapt to web-based classes well and, consequently, could be more susceptible to anxiety or even depression. Second, students with high levels of resilience might recover more quickly from adversities and cope with problems more actively, which could help with lowering their susceptibility to depression during the pandemic. Third, students who have high levels of resilience might be more likely to have successful experiences of coping with and recovering from adversities, including the pandemic.

The findings from this study showed that coping styles had a significant effect on relieving depression ($P<.001$), which is in line with previous studies [69,70]. Our study also showed that positive coping was inversely related to depression, while negative coping was positively related to depression among Chinese medical students. A positive coping style could help students cope with problems (eg, web-based classes) more rationally and might reduce stress, which in turn might prevent depression. This study also indicated that coping styles mediated the effect of resilience on depression. The coefficient of the resilience to depression path decreased after the coping style variables were added to the model, which indicated that coping styles had a partial mediating role in the relationship between resilience and depression. It is possible that individuals who are more resilient might be more likely to adopt a positive coping style, which might increase their likelihood of perceiving stressful situations, such as being quarantined at home while taking web-based classes during the COVID-19 pandemic, as surmountable. Therefore, they might be less likely to experience depression. Such individuals are also more likely to have better control over their emotions and be more motivated to figure out the solutions to problems, thus allowing them adapt to adverse situations. This could be beneficial for their mental well-being. Our results implied that positive coping and resilience training would be beneficial to medical students in terms of confronting the COVID-19 pandemic more positively and adapting to web-based classes more easily, which might help them to fight stress, reduce depression levels, and maintain mental well-being.

A few limitations exist in this study. First, this study is cross-sectional, which limits its ability to identify causal associations between variables. Second, self-reported measures may have resulted in response bias in this study. Third, the study sample only included medical students in years 1 to 3 from 1 university in northeastern China, which might limit the generalizability of our results.

Conclusion

This study found that Chinese medical students in web-based classes experienced slightly low levels of depression during the COVID-19 pandemic. Resilience ($P=.04$) and coping styles ($P<.001$) were significantly related to depression. Positive coping styles played an essential role in decreasing depression levels among medical students, while negative coping styles were positively related to depression. It was also found that coping styles mediated the association between resilience and depression. This study indicates that interventions that aim to enable the development of positive coping styles among individuals and improve their resilience are of great practical importance to decreasing depression levels among medical students while they are taking web-based classes during the large-scale COVID-19 pandemic.

Acknowledgments

All authors are very grateful to all participants who participated in this survey.

Conflicts of Interest

None declared.

References

1. Lin T, Stone SJ, Anderson T. Treating from afar: Mental health providers' challenges and concerns during the COVID-19 pandemic. Behav Med 2021 May 05:1-4. [doi: 10.1080/08964289.2021.1908217] [Medline: 33950782]
2. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
3. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
4. Education: From disruption to recovery. United Nations Educational, Scientific and Cultural Organization. URL: https://en.unesco.org/covid19/educationresponse [accessed 2021-05-25]

5. Ansert E, Rushing CJ. Feeling the burnout: perceptions of burnout, anxiety, depression, and personal achievement in US podiatric medical students. J Foot Ankle Surg. Epub ahead of print 2021 Mar 06. [doi: 10.1053/j.jfas.2021.02.007] [Medline: 33781641]
6. Shetty S, Shilpa C, Dey D, Kavya S. Academic crisis during COVID 19: Online classes, a panacea for imminent doctors. Indian J Otolaryngol Head Neck Surg 2020 Oct 17:1-5 [FREE Full text] [doi: 10.1007/s12070-020-02224-x] [Medline: 33102186]
7. Ibrahim AK, Kelly SJ, Adams CE, Glazebrook C. A systematic review of studies of depression prevalence in university students. J Psychiatr Res 2013 Mar;47(3):391-400. [doi: 10.1016/j.jpsychires.2012.11.015] [Medline: 23260171]
8. Naja WJ, Kansoun AH, Haddad RS. Prevalence of depression in medical students at the Lebanese University and exploring its correlation With facebook relevance: A questionnaire study. JMIR Res Protoc 2016 May 31;5(2):e96 [FREE Full text] [doi: 10.2196/resprot.4551] [Medline: 27246394]
9. Mayer FB, Santos IS, Silveira PSP, Lopes MHI, de Souza ARND, Campos EP, et al. Factors associated to depression and anxiety in medical students: a multicenter study. BMC Med Educ 2016 Oct 26;16(1):282 [FREE Full text] [doi: 10.1186/s12909-016-0791-1] [Medline: 27784316]
10. Dyrbye LN, Thomas MR, Shanafelt TD. Systematic review of depression, anxiety, and other indicators of psychological distress among U.S. and Canadian medical students. Acad Med 2006 Apr;81(4):354-373. [doi: 10.1097/00001888-200604000-00009] [Medline: 16565188]
11. Moutinho ILD, Maddalena NDCP, Roland RK, Lucchetti ALG, Tibiriçá SHC, Ezequiel ODS, et al. Depression, stress and anxiety in medical students: A cross-sectional comparison between students from different semesters. Rev Assoc Med Bras (1992) 2017 Jan 01;63(1):21-28 [FREE Full text] [doi: 10.1590/1806-9282.63.01.21] [Medline: 28225885]
12. Haldorsen H, Bak NH, Dissing A, Petersson B. Stress and symptoms of depression among medical students at the University of Copenhagen. Scand J Public Health 2014 Feb;42(1):89-95. [doi: 10.1177/1403494813503055] [Medline: 23999855]
13. Tang W, Hu T, Hu B, Jin C, Wang G, Xie C, et al. Prevalence and correlates of PTSD and depressive symptoms one month after the outbreak of the COVID-19 epidemic in a sample of home-quarantined Chinese university students. J Affect Disord 2020 Sep 01;274:1-7 [FREE Full text] [doi: 10.1016/j.jad.2020.05.009] [Medline: 32405111]
14. Wang ZH, Yang HL, Yang YQ, Liu D, Li ZH, Zhang XR, et al. Prevalence of anxiety and depression symptom, and the demands for psychological knowledge and interventions in college students during COVID-19 epidemic: A large cross-sectional study. J Affect Disord 2020 Oct 01;275:188-193 [FREE Full text] [doi: 10.1016/j.jad.2020.06.034] [Medline: 32734907]
15. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
16. Warwick HM, Salkovskis PM. Hypochondriasis. Behav Res Ther 1990;28(2):105-117. [doi: 10.1016/0005-7967(90)90023-c] [Medline: 2183757]
17. Rimes KA, Salkovskis PM. Prediction of psychological reactions to bone density screening for osteoporosis using a cognitive-behavioral model of health anxiety. Behav Res Ther 2002 Apr;40(4):359-381. [doi: 10.1016/s0005-7967(01)00015-8] [Medline: 12002895]
18. Taylor S. The Psychology of Pandemics: Preparing for the Next Global Outbreak of Infectious Disease. Newcastle upon Tyne, UK: Cambridge Scholars Publishing; 2019.
19. Goodwin R, Gaines Jr SO, Myers L, Neto F. Initial psychological responses to swine flu. Int J Behav Med 2011 Jun;18(2):88-92 [FREE Full text] [doi: 10.1007/s12529-010-9083-z] [Medline: 20195809]
20. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York, NY, USA: Springer Publishing Company; 1984.
21. Yang X, Yao L, Wu H, Wang Y, Liu L, Wang J, et al. Quality of life and its related factors in Chinese unemployed people: A population-based cross-sectional study. Int J Environ Res Public Health 2016 Aug 08;13(8):797 [FREE Full text] [doi: 10.3390/ijerph13080797] [Medline: 27509514]
22. Park N, Peterson C, Szvarca D, Molen RJV, Kim ES, Collon K. Positive psychology and physical health: Research and applications. Am J Lifestyle Med 2014 Sep 26;10(3):200-206 [FREE Full text] [doi: 10.1177/1559827614550277] [Medline: 30202275]
23. Seligman ME, Csikszentmihalyi M. Positive psychology. An introduction. Am Psychol 2000 Jan;55(1):5-14. [doi: 10.1037//0003-066x.55.1.5] [Medline: 11392865]
24. Sharpley CF, Bitsika V, Wootten AC, Christie DRH. Does resilience 'buffer' against depression in prostate cancer patients? A multi-site replication study. Eur J Cancer Care (Engl) 2014 Jul;23(4):545-552. [doi: 10.1111/ecc.12170] [Medline: 24506500]
25. Li J, Theng YL, Foo S. Does psychological resilience mediate the impact of social support on geriatric depression? An exploratory study among Chinese older adults in Singapore. Asian J Psychiatr 2015 Apr;14:22-27. [doi: 10.1016/j.ajp.2015.01.011] [Medline: 25703041]
26. Aburn G, Gott M, Hoare K. What is resilience? An Integrative Review of the empirical literature. J Adv Nurs 2016 May;72(5):980-1000. [doi: 10.1111/jan.12888] [Medline: 26748456]

27. Deng M, Pan Y, Zhou L, Chen X, Liu C, Huang X, et al. Resilience and cognitive function in patients with schizophrenia and bipolar disorder, and healthy controls. Front Psychiatry 2018 Jun 09;9:279 [FREE Full text] [doi: 10.3389/fpsyt.2018.00279] [Medline: 30008678]
28. Muñoz RF, Beardslee WR, Leykin Y. Major depression can be prevented. Am Psychol 2012;67(4):285-295 [FREE Full text] [doi: 10.1037/a0027666] [Medline: 22583342]
29. Weissberg RP, Kumpfer KL, Seligman MEP. Prevention that works for children and youth. An introduction. Am Psychol 2003;58(6-7):425-432. [doi: 10.1037/0003-066X.58.6-7.425] [Medline: 12971188]
30. Steinhardt M, Dolbier C. Evaluation of a resilience intervention to enhance coping strategies and protective factors and decrease symptomatology. J Am Coll Health 2008;56(4):445-453. [doi: 10.3200/JACH.56.44.445-454] [Medline: 18316290]
31. Jeste DV, Palmer BW. A call for a new positive psychiatry of ageing. Br J Psychiatry 2013 Feb;202:81-83. [doi: 10.1192/bjp.bp.112.110643] [Medline: 23377203]
32. Folkman S, Moskowitz JT. Coping: pitfalls and promise. Annu Rev Psychol 2004;55:745-774. [doi: 10.1146/annurev.psych.55.090902.141456] [Medline: 14744233]
33. Guardino CM, Schetter CD. Coping during pregnancy: a systematic review and recommendations. Health Psychol Rev 2014;8(1):70-94 [FREE Full text] [doi: 10.1080/17437199.2012.752659] [Medline: 24489596]
34. de Tychey C, Spitz E, Briançon S, Lighezzolo J, Girvan F, Rosati A, et al. Pre- and postnatal depression and coping: a comparative approach. J Affect Disord 2005 Apr;85(3):323-326. [doi: 10.1016/j.jad.2004.11.004] [Medline: 15780702]
35. Abdollahi A, Farab NM, Panahipour S, Allen KA. Academic hardiness as a moderator between evaluative concerns perfectionism and academic procrastination in students. J Genet Psychol 2020;181(5):365-374. [doi: 10.1080/00221325.2020.1783194] [Medline: 32578515]
36. Zheng Z, Han W, Li Y, Wang D, Gu S, Wang F. The mediating effect of coping style in the relationship between depression and disordered eating among Chinese female undergraduates. Front Psychol 2020 Jan 21;10:3011 [FREE Full text] [doi: 10.3389/fpsyg.2019.03011] [Medline: 32038388]
37. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
38. Chen S, Fang Y, Chiu H, Fan H, Jin T, Conwell Y. Validation of the nine-item Patient Health Questionnaire to screen for major depression in a Chinese primary care population. Asia Pac Psychiatry 2013 Jun;5(2):61-68. [doi: 10.1111/appy.12063] [Medline: 23857806]
39. Leung DYP, Mak YW, Leung SF, Chiang VCL, Loke AY. Measurement invariances of the PHQ-9 across gender and age groups in Chinese adolescents. Asia Pac Psychiatry 2020 Sep;12(3):e12381 [FREE Full text] [doi: 10.1111/appy.12381] [Medline: 32011101]
40. Zhang YL, Liang W, Chen ZM, Zhang HM, Zhang JH, Weng XQ, et al. Validity and reliability of Patient Health Questionnaire-9 and Patient Health Questionnaire-2 to screen for depression among college students in China. Asia Pac Psychiatry 2013 Dec;5(4):268-275. [doi: 10.1111/appy.12103] [Medline: 24123859]
41. Yang BX, Xia L, Huang R, Chen P, Luo D, Liu Q, et al. Relationship between eHealth literacy and psychological status during COVID-19 pandemic: A survey of Chinese residents. J Nurs Manag 2021 May;29(4):805-812 [FREE Full text] [doi: 10.1111/jonm.13221] [Medline: 33269492]
42. Block J, Kremen AM. IQ and ego-resiliency: conceptual and empirical connections and separateness. J Pers Soc Psychol 1996 Feb;70(2):349-361. [doi: 10.1037//0022-3514.70.2.349] [Medline: 8636887]
43. Windle G, Bennett KM, Noyes J. A methodological review of resilience measurement scales. Health Qual Life Outcomes 2011 Feb 04;9:8 [FREE Full text] [doi: 10.1186/1477-7525-9-8] [Medline: 21294858]
44. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
45. Woolston C. Signs of depression and anxiety soar among US graduate students during pandemic. Nature 2020 Sep;585(7823):147-148. [doi: 10.1038/d41586-020-02439-6] [Medline: 32811983]
46. Xiao H, Shu W, Li M, Li Z, Tao F, Wu X, et al. Social Distancing among Medical Students during the 2019 Coronavirus Disease Pandemic in China: Disease Awareness, Anxiety Disorder, Depression, and Behavioral Activities. Int J Environ Res Public Health 2020 Jul 14;17(14):5047 [FREE Full text] [doi: 10.3390/ijerph17145047] [Medline: 32674285]
47. Sun XJ, Niu GF, You ZQ, Zhou ZK, Tang Y. Gender, negative life events and coping on different stages of depression severity: A cross-sectional study among Chinese university students. J Affect Disord 2017 Feb;209:177-181. [doi: 10.1016/j.jad.2016.11.025] [Medline: 27923194]
48. Chen L, Wang L, Qiu XH, Yang XX, Qiao ZX, Yang YJ, et al. Depression among Chinese university students: prevalence and socio-demographic correlates. PLoS One 2013;8(3):e58379. [doi: 10.1371/journal.pone.0058379] [Medline: 23516468]
49. Song Y, Huang Y, Liu D, Kwan JSH, Zhang F, Sham PC, et al. Depression in college: depressive symptoms and personality factors in Beijing and Hong Kong college freshmen. Compr Psychiatry 2008;49(5):496-502. [doi: 10.1016/j.comppsych.2008.02.005] [Medline: 18702936]
50. Thompson G, McBride RB, Hosford CC, Halaas G. Resilience among medical students: The role of coping style and social support. Teach Learn Med 2016;28(2):174-182. [doi: 10.1080/10401334.2016.1146611] [Medline: 27064719]

51. Goebert D, Thompson D, Takeshita J, Beach C, Bryson P, Ephgrave K, et al. Depressive symptoms in medical students and residents: a multischool study. Acad Med 2009 Feb;84(2):236-241. [doi: 10.1097/ACM.0b013e31819391bb] [Medline: 19174678]
52. Mosley Jr TH, Perrin SG, Neral SM, Dubbert PM, Grothues CA, Pinto BM. Stress, coping, and well-being among third-year medical students. Acad Med 1994 Sep;69(9):765-767. [doi: 10.1097/00001888-199409000-00024] [Medline: 8074778]
53. Dahlin M, Joneborg N, Runeson B. Stress and depression among medical students: a cross-sectional study. Med Educ 2005 Jun;39(6):594-604. [doi: 10.1111/j.1365-2929.2005.02176.x] [Medline: 15910436]
54. Dyrbye LN, Thomas MR, Massie FS, Power DV, Eacker A, Harper W, et al. Burnout and suicidal ideation among U.S. medical students. Ann Intern Med 2008 Sep 02;149(5):334-341. [doi: 10.7326/0003-4819-149-5-200809020-00008] [Medline: 18765703]
55. O'Dowd E, O'Connor P, Lydon S, Mongan O, Connolly F, Diskin C, et al. Stress, coping, and psychological resilience among physicians. BMC Health Serv Res 2018 Sep 21;18(1):730 [FREE Full text] [doi: 10.1186/s12913-018-3541-8] [Medline: 30241524]
56. Conger RD, Conger KJ. Resilience in midwestern families: Selected findings from the first decade of a prospective, longitudinal study. J Marriage Fam 2002 May 02;64(2):361-373. [doi: 10.1111/j.1741-3737.2002.00361.x]
57. Werner EE. Risk, resilience, and recovery: Perspectives from the Kauai Longitudinal Study. Dev Psychopathol 1993;5(4):503-515. [doi: 10.1017/s095457940000612x]
58. Zhang J, Yang Z, Wang X, Li J, Dong L, Wang F, et al. The relationship between resilience, anxiety and depression among patients with mild symptoms of COVID-19 in China: A cross-sectional study. J Clin Nurs 2020 Nov;29(21-22):4020-4029 [FREE Full text] [doi: 10.1111/jocn.15425] [Medline: 32702192]
59. Dyrbye LN, Power DV, Massie FS, Eacker A, Harper W, Thomas MR, et al. Factors associated with resilience to and recovery from burnout: a prospective, multi-institutional study of US medical students. Med Educ 2010 Oct;44(10):1016-1026. [doi: 10.1111/j.1365-2923.2010.03754.x] [Medline: 20880371]
60. Kelifa MO, Yang Y, Herbert C, He Q, Wang P. Psychological resilience and current stressful events as potential mediators between adverse childhood experiences and depression among college students in Eritrea. Child Abuse Negl 2020 Aug;106:104480. [doi: 10.1016/j.chiabu.2020.104480] [Medline: 32470689]
61. Zhao F, Guo Y, Suhonen R, Leino-Kilpi H. Subjective well-being and its association with peer caring and resilience among nursing vs medical students: A questionnaire study. Nurse Educ Today 2016 Feb;37:108-113. [doi: 10.1016/j.nedt.2015.11.019] [Medline: 26694789]
62. Kim GM, Lim JY, Kim EJ, Park SM. Resilience of patients with chronic diseases: A systematic review. Health Soc Care Community 2019 Jul;27(4):797-807. [doi: 10.1111/hsc.12620] [Medline: 30027595]
63. Yu F, Raphael D, Mackay L, Smith M, King A. Personal and work-related factors associated with nurse resilience: A systematic review. Int J Nurs Stud 2019 May;93:129-140. [doi: 10.1016/j.ijnurstu.2019.02.014] [Medline: 30925279]
64. Bacchi S, Licinio J. Resilience and psychological distress in psychology and medical students. Acad Psychiatry 2017 Apr;41(2):185-188. [doi: 10.1007/s40596-016-0488-0] [Medline: 27060093]
65. Tempski P, Santos IS, Mayer FB, Enns SC, Perotta B, Paro HBMS, et al. Relationship among medical student resilience, educational environment and quality of life. PLoS One 2015 Jun 05;10(6):e0131535. [doi: 10.1371/journal.pone.0131535] [Medline: 26121357]
66. Ji G, Wei W, Yue KC, Li H, Shi LJ, Ma JD, et al. Effects of the COVID-19 pandemic on obsessive-compulsive symptoms among university students: Prospective cohort survey study. J Med Internet Res 2020 Sep 30;22(9):e21915 [FREE Full text] [doi: 10.2196/21915] [Medline: 32931444]
67. Kun P, Han S, Chen X, Yao L. Prevalence and risk factors for posttraumatic stress disorder: a cross-sectional study among survivors of the Wenchuan 2008 earthquake in China. Depress Anxiety 2009;26(12):1134-1140. [doi: 10.1002/da.20612] [Medline: 19842170]
68. Tugade MM, Fredrickson BL. Resilient individuals use positive emotions to bounce back from negative emotional experiences. J Pers Soc Psychol 2004 Feb;86(2):320-333 [FREE Full text] [doi: 10.1037/0022-3514.86.2.320] [Medline: 14769087]
69. Shao R, He P, Ling B, Tan L, Xu L, Hou Y, et al. Prevalence of depression and anxiety and correlations between depression, anxiety, family functioning, social support and coping styles among Chinese medical students. BMC Psychol 2020 Apr 22;8(1):38 [FREE Full text] [doi: 10.1186/s40359-020-00402-8] [Medline: 32321593]
70. Luo Y, Wang H. Correlation research on psychological health impact on nursing students against stress, coping way and social support. Nurse Educ Today 2009 Jan;29(1):5-8. [doi: 10.1016/j.nedt.2008.05.019] [Medline: 18692281]

Abbreviations

AMOS: Analysis of Moment Structures
BCa: bias-corrected and accelerated
CFI: comparative fit index
GFI: goodness-of-fit index

PHQ-9: Patient Health Questionnaire-9
RMSEA: root mean square error of approximation
TLI: Tucker-Lewis Index

Edited by G Eysenbach; submitted 25.10.20; peer-reviewed by XZ Hu, K Aguirre, Y Miao; comments to author 18.11.20; revised version received 09.12.20; accepted 19.05.21; published 07.06.21.

Please cite as:
Zhao L, Sznajder K, Cheng D, Wang S, Cui C, Yang X
Coping Styles for Mediating the Effect of Resilience on Depression Among Medical Students in Web-Based Classes During the COVID-19 Pandemic: Cross-sectional Questionnaire Study
J Med Internet Res 2021;23(6):e25259
URL: https://www.jmir.org/2021/6/e25259
doi:10.2196/25259
PMID:34033579

Original Paper

Patient Empowerment During the COVID-19 Pandemic by Ensuring Safe and Fast Communication of Test Results: Implementation and Performance of a Tracking System

Gunnar Völkel[1*], PhD; Axel Fürstberger[1*], PhD; Julian D Schwab[1*], PhD; Silke D Werle[1*], PhD; Nensi Ikonomi[1*], MSc; Thomas Gscheidmeier[2], MSc; Johann M Kraus[1], PhD; Alexander Groß[1], PhD; Martin Holderried[3], MD, PhD; Julien Balig[1], MSc; Franz Jobst[4], MSc; Peter Kuhn[5], PhD; Klaus A Kuhn[6], PhD; Oliver Kohlbacher[7], PhD; Udo X Kaisers[4], MD; Thomas Seufferlein[8*], MD; Hans A Kestler[1*], PhD

[1]Institute of Medical Systems Biology, Ulm University, Ulm, Germany
[2]Department of Clinical Chemistry, University Hospital Ulm, Ulm, Germany
[3]Department of Medical Development and Quality Management, University Hospital Tübingen, Tübingen, Germany
[4]University Hospital Ulm, Ulm, Germany
[5]Comprehensive Cancer Center, University Hospital Ulm, Ulm, Germany
[6]Institute of Medical Informatics, Statistics and Epidemiology, Technical University of Munich, Ulm, Germany
[7]Institute for Translational Bioinformatics, University Hospital Tübingen, Tübingen, Germany
[8]Department of Internal Medicine I, University Hospital Ulm, Ulm, Germany
[*]these authors contributed equally

Corresponding Author:
Hans A Kestler, PhD
Institute of Medical Systems Biology
Ulm University
Albert-Einstein-Allee 11
Ulm, 89081
Germany
Phone: 49 731 500 24500
Fax: 49 731 500 24501
Email: hans.kestler@uni-ulm.de

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/6/e31253/

Abstract

Background: Overcoming the COVID-19 crisis requires new ideas and strategies for online communication of personal medical information and patient empowerment. Rapid testing of a large number of subjects is essential for monitoring and delaying the spread of SARS-CoV-2 in order to mitigate the pandemic’s consequences. People who do not know that they are infected may not stay in quarantine and, thus, risk infecting others. Unfortunately, the massive number of COVID-19 tests performed is challenging for both laboratories and the units that conduct throat swabs and communicate the results.

Objective: The goal of this study was to reduce the communication burden for health care professionals. We developed a secure and easy-to-use tracking system to report COVID-19 test results online that is simple to understand for the tested subjects as soon as these results become available. Instead of personal calls, the system updates the status and the results of the tests automatically. This aims to reduce the delay when informing testees about their results and, consequently, to slow down the virus spread.

Methods: The application in this study draws on an existing tracking tool. With this open-source and browser-based online tracking system, we aim to minimize the time required to inform the tested person and the testing units (eg, hospitals or the public health care system). The system can be integrated into the clinical workflow with very modest effort and avoids excessive load to telephone hotlines.

Results: The test statuses and results are published on a secured webpage, enabling regular status checks by patients; status checks are performed without the use of smartphones, which has some importance, as smartphone usage diminishes with age. Stress tests and statistics show the performance of our software. CTest is currently running at two university hospitals in Germany—University Hospital Ulm and University Hospital Tübingen—with thousands of tests being performed each week. Results show a mean number of 10 (SD 2.8) views per testee.

Conclusions: CTest runs independently of existing infrastructures, aims at straightforward integration, and aims for the safe transmission of information. The system is easy to use for testees. QR (Quick Response) code links allow for quick access to the test results. The mean number of views per entry indicates a reduced amount of time for both health care professionals and testees. The system is quite generic and can be extended and adapted to other communication tasks.

(J Med Internet Res 2021;23(6):e27348) doi:10.2196/27348

KEYWORDS

process optimization; patient empowerment; data security; COVID-19; clinical information system; platform independent; eHealth; telemedicine; quality management

Introduction

After the first outbreak of SARS-CoV-2 in December 2019 in Wuhan, Hubei province, China, the virus spread rapidly worldwide [1,2]. Thus, the management of its induced crisis has become a ubiquitous topic [3]. Each day, the number of new infections increases worldwide, and reached 107,021,165 confirmed cases globally on February 10, 2021, at 4:22 PM [4].

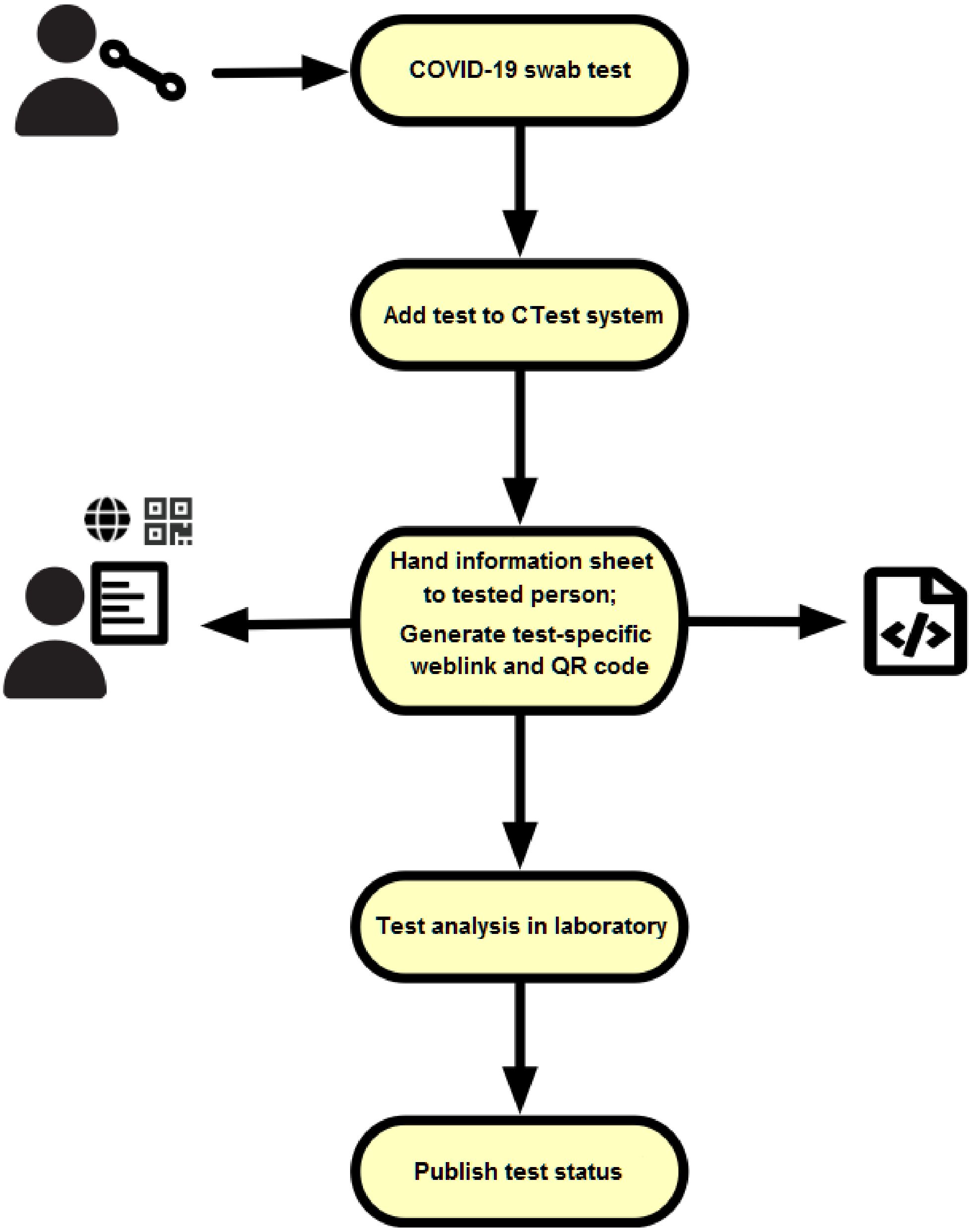
The current pandemic has significantly affected the possibility of direct interpersonal communication, together with a massive overload of the public health care system. In this context, digital technologies have become crucial sources of support. While surveys and data collection have empowered the evaluation of the first lockdown measures [5], the development of apps and dashboards is paramount in controlling the virus's spread [6]. Here, rapid case identification is one of the demanding tasks in controlling its spread. The most common available test for COVID-19 infections is to take a throat swab and test by real-time reverse transcription–polymerase chain reaction (RT-PCR) [7-9]. While speeding up this diagnostic test itself has limited feasibility, the time lag between administration of the test and the communication of results can be improved to ensure adequate isolation of positive cases. For ensuring maximal containment of spread, different digital solutions have been applied: (1) rapid identification of cases has been supported by apps concerning contact tracing and tracking of self-reported symptoms [10-14] and (2) special attention has been paid to artificial intelligence approaches supporting home-based self-testing and diagnosis [10,14]. In the case of Germany, available official apps to support health management have been developed and are provided in Table 1 [15-25]. Nevertheless, given the lack of standardization of home-based testing and self-made diagnosis [26,27], empowering the communication between testees and public health institutions is crucial in overcoming the current pandemic crisis. Here, the natural delay between hospital-based tests and the resulting communication with testees has not yet been addressed. Timely communication of results is essential for taking appropriate action; however, it is challenging, due to the high throughput and the high demand under the circumstances in which the tests are carried out. In most European countries, the current practice is to call the respective clinic or laboratory for information about test results [28]. This process ties up considerable resources and does not scale well for large numbers of tests [28]. Communication channels become overloaded and time-consuming for medical staff due to the testees' repeated calls, as waiting times and anxiety of the testees increases, altogether affecting patient empowerment.

Table 1. Official COVID-19–associated apps released in Germany.

App	Institution of release in Germany	Function
Corona-Warn-App [15]	Federal Government, Robert Koch Institute	Anonymized contact tracking Incidence and case reporting Personalized contact tracking
Corona-Datenspende [16]	Robert Koch Institute	Collection of provided health data from fitness trackers and smartwatches
CovApp [17]	Charité and Data4Life	Anonymized questionnaire providing contacts and guidance
Warn-App NINA [18]	Federal Department of Civil Protection and Disaster Assistance	Basic information about the COVID-19 crisis Local valid regulations
CoCoV app [19]	University Hospital Ulm	Private diary on the tolerability of COVID-19 vaccination based on a standardized questionnaire Aggregated analysis and presentation of events that have occurred on a public dashboard
Corona Check [20]	University Hospital Würzburg	Self-screening of symptoms with recommendations for action based on the results Live ticker with official regulations Recommendations on how to protect oneself and others
Corona Health app [21]	University Hospital Würzburg	Anonymized questionnaires on the psychological and physical health-related impacts of COVID-19 Screening analysis News
Safe Vac 2.0 [22]	Paul-Ehrlich-Institut	Anonymized tracking of tolerability of COVID-19 vaccination
WHO Academy: Covid-19 Learning [23]	World Health Organization	Guidance for case management, infection prevention control, and laboratory testing Epidemiology Regional and international information and regulations Research and development
Gutenberg COVID-19 Studien app [24]	University Medical Center Mainz	Survey of study participants
CoronaBoXX [25]	Charité	Health documentation of people in quarantine Hygiene checklist

We set up a framework to ensure fast and secure communication of health-related results to testees. This was applied to accelerate the COVID-19 testing procedure. Therefore, we implemented an online query system, called CTest, that provides convenient digital access to the status and results of the respective COVID-19 tests to testees (Figure 1). This approach avoids unnecessary and repeated phone calls, avoids manual transcription errors, limits language-associated barriers, and, consequently, reduces the burden on the clinical staff. CTest empowers testees, since they can check their test results and statuses independently through a secure online system [29].

Figure 1. Workflow of the COVID-19 test process using CTest. First, a COVID-19 swab test is performed. The test is then added to the CTest system via an order number. Based on this order number, CTest generates a cryptographically secure tracking ID. An information sheet with a test-specific weblink and QR (Quick Response) code is generated by CTest and handed over to the person tested. After laboratory analysis is complete, the results are sent to CTest. CTest updates the test results, which can be queried via the individual weblink.



The CTest system extends the functionalities of the previously developed online tracking tool TraqBio [30]. We created this application to simplify and standardize communication between users and core facilities. Its clean design and its open-source license allow for the development (ie, refactoring), deployment, establishment, and integration of CTest within a short period.

CTest runs on Java Virtual Machine [30,31]. The web application can be deployed independently of the operating system and platform.

Methods

Implementation and Setup

We based the CTest application on the TraqBio software [30]. The back-end functionality of the CTest server was implemented in the Lisp dialect Clojure. Clojure runs on Java Virtual Machine (Oracle) [31] and is, thus, platform independent. The back end comprises a database that stores the scheduled COVID-19 tests and corresponding results. Clojure supports different databases and connectors. We chose to use SQLite, version 3.29.0 [32], to keep the setup independent of additional database servers. A local database file holds the data. Additionally, CTest provides functionality for user management and data backup.

Every new order number entered and supplied by the laboratory staff in the CTest system generates a database entry with a unique identifier and unique tracking number. Here, this order number is used as the primary unambiguous identifier. CTest can be configured to append the current date to the order number to create a primary unambiguous identifier in setups where order numbers are only guaranteed to be unique within the same day. Using a secure random number generator, we generate a corresponding tracking number for each new entry [33]. This random number generator generates 6 bytes, which are then translated into a sequence of 12 characters containing numbers from 0 to 9 and letters from A to F. This sequence of characters is unique and is not created in sequential order. Every tracking number generated is checked for uniqueness prior to use (ie, there is no link between the tracking number and the order number).

Web-based front-end functionality is implemented using JavaScript and HTML templates. We used freely available standard web frameworks, such as Bootstrap v3 (Twitter) [34,35], jQuery (OpenJS Foundation) [36], and extensions to these frameworks. Using these state-of-the-art frameworks, we aim to enable a straightforward adaptation of the front end for integration at other institutions. The front end features a responsive graphical user interface for the management functions, the creation of tests, and accessibility for users to their test results. Persons may query their test status and results via a unique weblink without requiring an account or log-in. As an alternative to the weblink, a QR (Quick Response) code is provided for easy access.

To secure web communication, creation of new entries, and queuing of test results, our setup consists of different security layers. Only Secure Sockets Layer (SSL)–certified access to the websites is allowed via HTTP Secure (HTTPS). Connection and transferred data are encrypted, and the certificate authority–signed server certificates are used. A reverse-proxy setup forwards the external hostname to a virtual machine within the hospital's secure network infrastructure. On the virtual server, another reverse proxy is in place to allow for running of the Java application as a nonprivileged user. As the standard HTTP port 80 is privileged, it can only be used by a system user, so running the application as a user with system-wide rights is a security risk. Therefore, proxy settings forward the privileged port to a high port (above 1024) that normal users can control. Hence, no system rights are necessary for managing the application at the operating system level. For the clinical environment setup, we also set up a firewall with specific ban rules—iptables and fail2ban service on a Linux operating system—to prevent brute force attacks on log-in or tracking numbers. Also, network Internet Protocols (IPs) and subnets can be white-listed to allow access to management functions of the application. Thus, other computers and devices are blocked from accessing these functions after failed attempts. For the tracking interface, a brute force attack, such as trying all combinations of possible tracking numbers, is shielded by blocking IPs after too many failed attempts with wrong or nonexistent tracking numbers.

Query Performance Test

We first performed 1024 simultaneous requests for our performance test using a single machine and a single network connection. Next, we created a mixed data set (ie, interleaved_urls), including 758 available database entries (ie, available_urls) and the same number of unavailable tracking numbers (ie, notavailable_urls). Furthermore, we measured repeated queries of a single available URL, repeated queries of a single unavailable URL, as well as repeats of these two URLs consecutively. That makes six data sets in total, each of which was accessed 1516 times. Queries were carried out once in sequence (ie, ordered) and randomly (ie, random). Also, we tested the two-scenario caching function of the browser (ie, 1filePerRequest) and complete reloading (ie, 23or1filesPerRequest). This means that unchanged files are not reloaded (caching), whereas in the other scenario, all required files and displayed images (eg, flag graphics or Cascading Style Sheets) are reloaded.

For the second scenario, one file is loaded if the tracking number is not available and 23 files are loaded if the tracking number is available. Stress tests were measured using Siege 4.0.4 (JoeDog Software) [37].

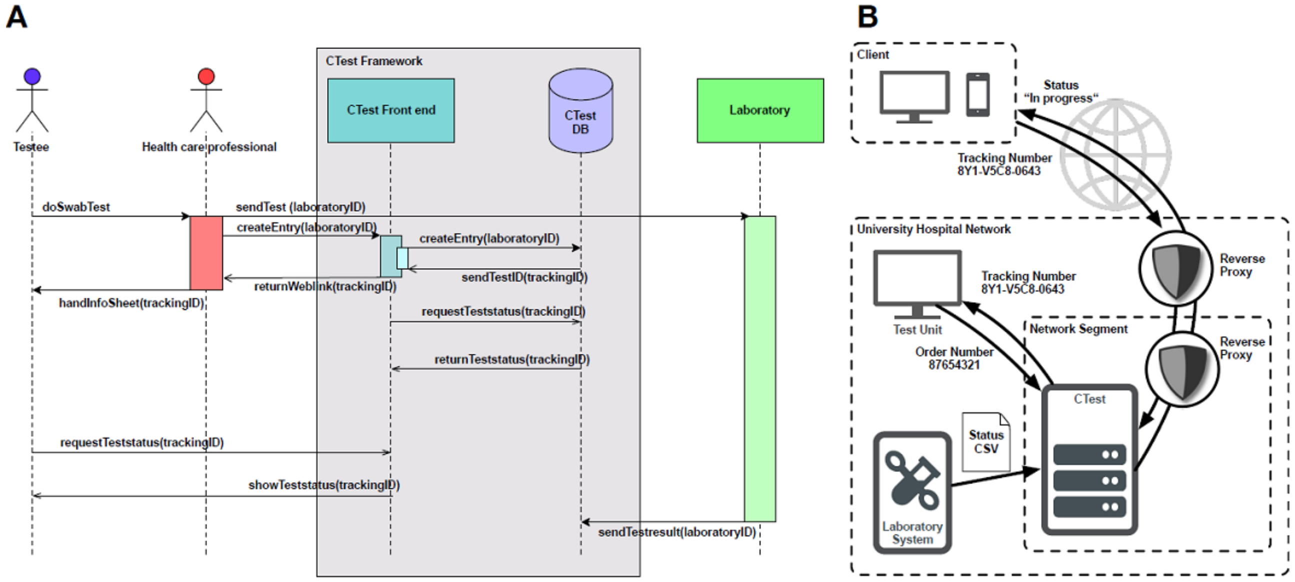
Data Availability

The source code, documentation, and an installation guide are freely available from GitHub [38] under the Eclipse Public License v2.0.

Results

CTest was built based on an existing, proven software stack: TraqBio [30]. It extends TraqBio to the functionality required for COVID-19 tests. We were able to successfully integrate it into the existing clinical testing workflows for SARS-CoV-2 infections in two major German university medical centers—University Hospital Ulm and University Hospital Tübingen—within a few days (Figure 2).

Figure 2. COVID-19 test procedure and setup. (A) The left-hand panel shows the detailed process flow of the COVID-19 tests. CTest (gray) is tightly integrated into the test process. The test framework comprises a web-based front end (turquoise) and a Clojure back end, including a mySQL (Structured Query Language) database (DB) (purple). After the swab test for COVID-19, the health care professional creates a CTest entry via entering the laboratory ID in the CTest front end. Based on that, a secure tracking ID is created by the back end and a corresponding personalized URL is created. CTest then automatically creates a test-specific information sheet for each testee. This sheet contains the test-specific weblink and a QR (Quick Response) code, both based on the tracking ID. The URL does not provide any information to track back the test results to the corresponding testee (blue). Via this URL, a testee can request the corresponding test status at any time. After the analysis in the laboratory, results are transmitted to the CTest database. The CTest front end automatically updates the status on the test-specific weblink based on the corresponding test results. (B) In the right-hand panel, a possible integration scenario of CTest into a clinical infrastructure is shown. CTest runs on a virtual machine within the secured network of a university hospital. The external hostname, with weblinks to the testee's status, is forwarded via reverse proxy. A second reverse proxy forwards the port to the application to a nonprivileged port. Thus, the application does not need to run with a system user. Inside the hospital's network, the test unit communicates to the CTest server to create new test cases via an order number. The CTest server returns back the tracking number and the corresponding weblink and QR code. The laboratory pushes test results as comma-separated values (CSV) files to the CTest server. The test-specific webpage content is then updated according to the test result.



The workflow starts with taking a sample for testing. In the first step, an order number is generated by the testing lab and added as a barcode label to the sample. Analogous to other medical applications, order numbers within the clinical or laboratory information system are unique but do not contain personal information about the patient [39]. This order number is transferred to the CTest system via scanning or typing the number into the dialogue window’s input field. We implemented format restrictions via regular expressions to the input field to minimize incorrect entries. Afterward, CTest generates an unambiguous, nonconsecutive tracking number. Therefore, a cryptographically strong pseudo-random number generator [33] creates 6 bytes that are transferred into a 12-digit character code, including letters from A to F and numbers from 0 to 9. Using this tracking number ensures that no personal information about any testee can be inferred.

After taking the sample, the tracking number is given to each testee on a printed sheet, including information on how to access the status of their COVID-19 test (Figure 3, A). After the sample has been processed, the lab system sends updated files as comma-separated values (CSV) files via an encrypted Secure File Transfer Protocol (SFTP) connection to the CTest server. Results in the CSV file are then automatically parsed, backed up, and imported into the CTest database, which leads to an automatic status update of each processed test.

Figure 3. Information for testees. Information concerning access to the test status and its results are given in eight languages. (A) The information sheet handed over to testees informs them of how to access the COVID-19 test results online. (B) The test-specific URL links to a page that includes information about the test result (eg, negative test result) or (C) information stating that the test is still in progress. (D) Scanning the QR (Quick Response) code with a smartphone link takes the testee directly to the current test status or its result.

ANR 23434234-20200323

Um Ihnen einen einfachen und möglichst schnellen Zugriff auf den Status Ihres Tests zu geben, haben wir einen Internetdienst eingerichtet. Sie haben direkten Zugriff auf den Status Ihres Tests über folgende Internetadresse:

To provide you a quick and easy access to your test result we set up a dedicated internet service. You can reach your personal test result using following website:

Pour vous offrir un service simple et rapid pour l'accès votre test, nous avons ouvré un service direct via l'adresse Internet suivante:

Testinizin sonucuna kolay ve hızlı bir şekilde erişebilmeniz için bir internet hizmeti kurduk. Aşağıdaki internet adresi aracılığıyla testinizin sonucuna doğrudan erişebilirsiniz:

Al fine di fornire un veloce e facile accesso al risultato del test, abbiamo messo a punto un servizio internet dedicato. É possibile avere accesso al proprio risultato utilizzando il seguente sito:

Мы устроили интернет-сервис, чтобы вы могли просто и быстро получить доступ к статусу вашего теста. Прямой доступ к статусу вашего теста по следующему интернет-адресу:

为了快速方便地访问您的测试状态，我们实施了一项互联网服务。 您可以通过以下网站直接访问测试状态。：

لقد أنشأنا لك خدمة إنترنت لتمنحك وصولاً سهلاً وسريعًا إلى حالة و نتيجة الاختبار. لديك امكانية الوصول المباشر إلى حالة الاختبار الخاص بك عبر عنوان الإنترنت التالي:

https://ctest.server.de/track/48Y1-V5C8-0643

Tracking Nr.
48Y1-V5C8-0643

Status

Test negativ
Hinweis: ein negatives Ergebnis schließt die Möglichkeit einer Coronavirus-Infektion nicht vollständig aus (z.B. bei Probenabnahme zu früh während des Infektionsverlaufs).

Result negative
Information: a negative test result does not entirely exclude a coronavirus infection (e.g. in case samples were obtained too early after infection).

Test négatif
Remarque: un résultat négatif n'éxclut pas entièrement la possibilité d'une infection avec le coronavirus (par example, si des échantillons sont prélevés trop tôt au cours de l'infection).

Negatif test
Not: negatif bir sonuç, koronavirüs enfeksiyonu olasılığını tamamen dışlamaz (örneğin, enfeksiyon sırasında örnek çok erken alınırsa).

Test negativo
Nota: un risultato negativo non esclude completamente la possibilità di un'infezione da coronavirus (ad es. Se il campione viene prelevato troppo presto nel corso dell'infezione).

Тест отрицательный
Примечание: отрицательный результат не полностью исключает возможность коронавирусной инфекции (например, если образец был взят слишком рано в течение инфекции).

测试结果阴性
注意：阴性结果并未完全排除冠状病毒感染的可能性（例如，在感染过程中过早取样）。

النتيجة سلبية
ملاحظة: النتيجة السلبية لا تلغي تمامًا احتمالية الإصابة بالفيروس كورونا (على سبيل المثال إذا تم أخذ العينة في وقت مبكر جدًا من تقدم الإصابة)

Tracking Nr.
48Y1-V5C8-0643

Status

Analyse läuft
Test in progress
Analyse en cours
Analiz devam ediyor
Analisi in corso
Анализ в обработки
进行中的分析
ولا يزال التحليل مستمرة

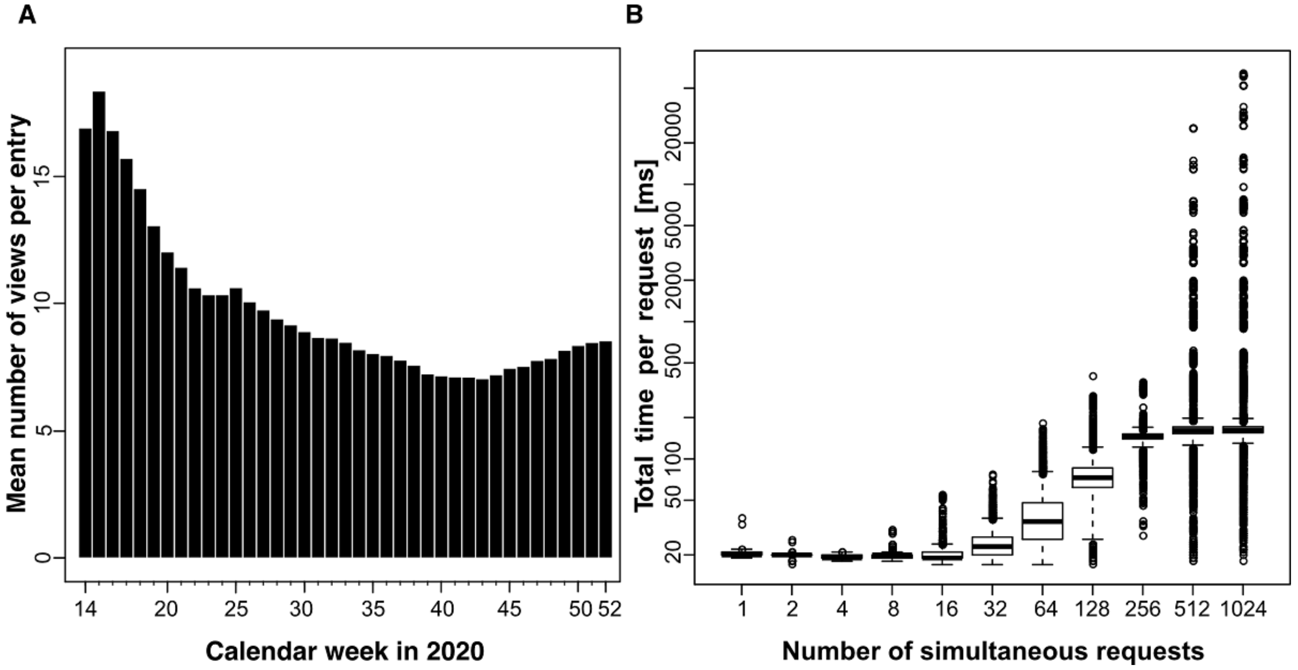
Currently, the CTest system distinguishes multiple potential outcomes: (1) the COVID-19 test is negative (Figure 3, B), (2) the COVID-19 test is still in progress (Figure 3, C), (3) the COVID-19 test is positive, or (4) the COVID-19 test is flawed. As the software is generic, one can implement other outcomes. Phone calls from the health department still inform people who have tested positive for COVID-19. Thereby, officials can inform them about health arrangements and how to avoid further spreading of the virus.

We provide two possibilities to query the status of the COVID-19 test. Testees can either scan a QR code on the information letter they received with a smartphone and get redirected to a test status webpage (Figure 3, D) or they can enter the weblink from the information letter into a web browser directly. The status of the individual test result is automatically displayed in a responsive form on the device (Figure 3, D). We have observed that both possibilities for requesting test results have been used. A total of 30.6% (64,201/209,808) of people preferred to query the test results via a web browser, while the rest of the testees preferred to scan the provided QR code.

To overcome language barriers, we translated information concerning the procedure to obtain test results, and the results themselves, into eight languages: German, English, French, Turkish, Italian, Russian, Chinese, and Arabic. Here, native-language speakers were asked to translate the information and kindly supported the translation process to ensure correct translations. Additionally, in the context of accessibility for medical information and patient empowerment, we paid particular attention to the use of easily understandable language and integration of behavioral recommendations.

Highlighting the reduced burden for clinical staff during the pandemic and the potential scalability of this approach in general, we estimate that there will be a personnel expense according to the derived data. On the one hand, we assume a mean of 10 (SD 2.8) views per entry, as shown in our data (Figure 4, A), and a duration of 30 seconds per access event via CTest. These numbers result in a total time consumption of 5 minutes via the use of CTest. On the other hand, a telephone inquiry is estimated to take 3 minutes if the lines are not busy. Using the number of times that CTest was accessed as the number of phone calls, the estimated amount of time per entry increases to 30 minutes. However, as a more realistic approach, we assume three calls per entry per day, which, at best, would lead to around 9 minutes of inquiry time for the patient.

Figure 4. Requests to the CTest server. (A) Since its introduction into the routine at the University Hospital Ulm (Germany) in week 14 (2020), a mean of about 10 (SD 2.8) views per test entry was recorded. The plot shows the mean number of views per entry per calendar week in 2020 as bars. (B) The server can respond within 200 ms to >80% of requests when responding to up to 1024 simultaneous requests. In this simulation, requests were made using a single machine and a single network connection.



Furthermore, we checked its performance and robustness in load tests (Figure 4, B, and Figures S1-S6 in Multimedia Appendix 1). Our CTest server can respond within 200 milliseconds to over 80% of requests and within 500 milliseconds to over 90% of requests when responding to up to 1024 simultaneous requests (Figure 4, B). Additionally, we performed stress tests with available and unavailable tracking numbers (Figures S1-S6 in Multimedia Appendix 1). Based on these tests, we are confident that CTest is well-suited to rapid testing demands even if deployed in an ad hoc manner on standard hardware. Its platform independence allows for its deployment on a wide variety of existing infrastructures.

Due to the high number of tests being administered during the peak times of COVID-19 infection waves, CTest was designed to handle large numbers of queries in a short amount of time. At Ulm University Hospital, a maximum of 394 laboratory tests were added to CTest per day. In total, 24,643 tests were added to CTest in 2020. In contrast, we registered up to 5263 access events to the website within one day and 209,808 in total in 2020. Since the introduction of CTest into the clinical routine, there have been around 12 queries performed per test (Figure 4, A). The handling of all these requests via phone calls would lead to communication overload. Consequently, CTest can reduce the burden on clinical staff. On top of that, CTest data at Ulm University Hospital shows that test results are provided to testees within one day.

Another feature of CTest is its functionality for error reporting and statistics. A dedicated *reporter* account is required to access the reporting data. Here, all data are provided in the machine-readable JavaScript Object Notation (JSON) format and the path “/reports/list” can be accessed to get a list of information and error reports. The included information is about the successful backup runs and successful test status imports from the CSV files. Also, the number of test result views per day can be accessed at “/reports/views.” General system information, such as memory consumption and central processing unit usage, is available at “/reports/system.” Furthermore, the sample collection dates are available at “/reports/test-dates” for analysis.

We present CTest as a web interface. In addition to its web interface, CTest facilitates third-party software (ie, apps) access if the testee decides to use it. These apps can query the status using the tracking link with appended “?app=true,” which returns only the test status instead of the complete HTML document.

Discussion

Fast and efficient communication with patients is the key to effective treatments and avoidance of health-related misunderstandings [40-42]. The COVID-19 crisis has brought a greater level of attention to this topic. In particular, hospital overload and lack of possibilities for direct and in-person communication have highlighted huge limitations for patient empowerment and compliance. On these grounds, we developed a framework to reduce the communication burden and to empower patients. The CTest framework enables an automated system that is always available and that does not require continuous human supervision and office hours. This framework thus reduces the burden on clinical staff involved in the COVID-19 crisis and slows down the spread of the virus by quickly and easily informing people who have been tested with concrete recommendations for action. Furthermore, we wanted to empower testees to obtain their results in a facile and easy-to-access way while, at the same time, ensuring efficient and almost instantaneous and exclusive communication. In the same way, other applications have tried to address patient empowerment and rapid isolation of positive cases [10-14]. Here, home-based self-tests and diagnoses, together with contact tracking, are major pillars that attempt to relieve the burden on public health. These approaches are meant to reduce the number of people with potential COVID-19 infections coming to hospitals for testing and to empower patients by providing insights into their health status [6]. Nevertheless, they still lack reliability and sensitivity [6]. For this reason, it is of relevance to support the information process in the public health context. In general, both aspects together are crucial in the effort to contain the spread of COVID-19. With our approach, we provide a unique workflow to support communication. Even if our implementation attempts to speed up hospital-related processing, we are aware that other steps, from testing to communicating results, are affected by delays. Nevertheless, the presented approach has the potential to support multiple testing locations, therefore, synergistically complementing the effort to provide reliable testing for patients. In this context, our results encourage the application of CTest.

CTest has already been successfully integrated into the hospital information system and captures thousands of COVID-19 tests per week at the University Hospital Ulm and at the University Hospital Tübingen in Germany. As part of the clinical routine, the first analyses of CTest showed a mean of 10 queries per test performed. Even half of that number of telephone inquiries would lead to communication overload. Consequently, the introduction of CTest into the clinical routine could achieve our primary goal of reducing the burden on clinical staff. The open-source license of TraqBio and its clean and simple setup were beneficial for this purpose [30]. This made it possible to implement and integrate the CTest system within a very short period (4 days) into the clinical workflow. Another advantage is that users do not have to create an account to request their test results. Besides, we provide two possibilities for querying the status of the test results. The fact that 30.6% of users queried their test results via QR codes encourages us in this implementation.

Our workload tests demonstrated the ability of CTest to deal with a massive number of parallel access events. These could not be processed via telephone calls. With the numbers estimated in our Results section, one employee could handle around 50 phone calls per 8-hour shift. This fact highlights the potential for increasing the speed of informing testees and for slowing down the virus spread. Using CTest, test results at Ulm University Hospital are available within a mean of one day. In contrast to that, for instance, Omar et al [43] imputed an average delay of 5.56 days in notifying testees with COVID-19–positive results using the conventional way. The estimated results from this study emphasize the reduced amount of time for testees to receive results regardless of business hours and, accordingly, the increasing potential for self-empowerment. Additionally, modeling approaches highlighted the relevance of rapid testing in flattening the infection curve [44,45].

Reducing public health burden is crucial in the final aim of providing digital health support for patient empowerment. In fact, the first consequence for patients of a rapid test response is an overall effect on their stress levels. After being tested, people are worried until they know the result of their test [46]. Notably, in the oncological context, it was shown that stress levels of patients waiting for their diagnosis were comparable to the ones of patients receiving severe diagnostic results [47]. Moreover, not knowing the outcome might tempt testees not to act according to recommendations and, thus, increase the risk of infecting others. Another crucial point we wanted to address with our workflow is ensuring comprehensible and secure communication. For this reason, we implemented translations in eight different languages in CTest to avoid language barriers as much as possible. To ensure the correct understanding of the provided information, we contacted native-language speakers who kindly provided translations. Previous studies outside of the pandemic context already showed that nonnative speakers experienced more stress and uncertainty in communicating with local health care providers. Steinberg and colleagues [42] reported a case study on safe and high-quality health care for children whose parents have limited English proficiencies in the United States. By performing 48 interviews with Latina mothers from two independent cities, they were able to show that these mothers experienced frustration with health care providers. Similar results were obtained from a South African study from Hunter-Adams and Rother [41]. Here, they investigated language barriers between local health care providers and cross-border migrants. Again, interviewed subjects reported frustration in understanding health care indications. Strikingly, some of the participants reported invasive medical procedures performed without explicit consent. Overall, not understanding indications increases fears over unwanted procedures and inaccessibility of health care, which, in the end, affects patient compliance. In the context of the COVID-19 pandemic, these issues would lead to increased communication overload and, consequently, to uncontrolled viral spread. For this reason, we tried to maximize the availability of translation options. The presented workflow also limits the occurrence of third-party translation that would affect patient privacy. In addition, it was shown that partners that try to play the role of interpreter caused additional nonprofessional nonmedical interpretations, again affecting patient empowerment and compliance [41].

In the interest of privacy protection, tracking numbers for test results in our workflow are created based on nonpersonalized order numbers. Based on this implementation, we address big data challenges in personalized medicine [29] and respect the German and European data protection laws. To tackle these data protection issues, clinical or laboratory information systems are often closed-source systems. Therefore, development or integration of new interfaces can be time-consuming. To overcome these barriers, we developed an independent, stand-alone software solution without storing personalized data. Nevertheless, this has the limitation wherein CTest alone cannot automatically transfer positive case results to the public health department. Since this transfer is compulsory in Germany, additional communication still has to be made by employees. In addition, a required step is the input of test results from external sources. Consequently, CTest is independent of existing infrastructures, such as specific laboratory information systems. Data import interfaces can be adapted for a broad range of scenarios (eg, hospitals, independent test centers, or resident physicians).

Currently, CTest is specialized for the query of COVID-19 test results and their status. However, it is a generic framework that can readily be adapted to other queries or to the distribution of different types of test results in all medical fields. Even integrating the CTest system into apps is possible if the HTML view presented here is not desired. For this purpose, the token “?app=true” has to be added after the tracking number. All in all, the idea of CTest is a generalizable approach that can be adapted to various uses in medical communication, such as blood analysis for resident physicians. In this context, an outlook for further development of the software is to have a user interface for configuration. Respective requirements for results, the connected database, and other parameters could be set via a graphic user interface to ease a roll-out into other domains. This might become interesting in the context of electronic prior authorization processes.

Acknowledgments

We thank the COVID-19 task force of the University Hospital Ulm, the Department of Clinical Chemistry, the Comprehensive Cancer Center Ulm, the Zentrum für Information und Kommunikation (ZIK), and all our supporters at University Medicine Ulm—Martin Loose, Thomas Baur, Robert Mahnke, and others—who supported the development of the tool and allowed its rapid integration into the clinical processes. We also thank Yessin Yahia, William Ferdinand, Cagatay Günes, and Xenia Kirgis for their help in translating the texts. The rapid implementation, deployment, and integration of the application were also made possible by support from different digitization initiatives.

KK, HK, and OK acknowledge funding from the Germany Federal Ministry of Education and Research (BMBF) as part of the DIFUTURE project, Medical Informatics Initiative (grants 01ZZ1804I and 01ZZ1804D). OK and HK acknowledge funding from the Ministry of Science and Art Baden-Württemberg (Zentrum für Innovative Versorgung [ZIV]). OK and TS acknowledge funding from the Ministry of Social Affairs of the state of Baden-Württemberg (Zentren für Personalisierte Medizin [ZPM]). HK acknowledges funding from the German Science Foundation (DFG) (grant 217328187). HK and UK acknowledge funding from the German BMBF (Nationales Forschungs Netzwerk [NFN] COVID-19) (grant 01KX2021).

Authors' Contributions

KK, OK, PK, UK, TS, and HK were responsible for funding acquisition. HK was responsible for project administration. GV, AF, TG, JK, AG, and JS were responsible for software development. HK was responsible for supervision of the project. GV, AF, JS, and SW were responsible for visualization of the results and for the development of software graphics. JS, SW, AF, PK, OK, GV, NI, and HK were responsible for writing the original draft of the manuscript. JS, GV, KK, SW, NI, JB, FJ, MH, OK, TS, and HK were responsible for writing, reviewing, and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Performance and robustness tests.
[PDF File (Adobe PDF File), 5992 KB - jmir_v23i6e27348_app1.pdf]

References

1. Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. J Med Virol 2020 Apr;92(4):441-447 [FREE Full text] [doi: 10.1002/jmv.25689] [Medline: 31994742]
2. Zhu H, Wei L, Niu P. The novel coronavirus outbreak in Wuhan, China. Glob Health Res Policy 2020;5:6 [FREE Full text] [doi: 10.1186/s41256-020-00135-6] [Medline: 32226823]
3. Cohen J, Kupferschmidt K. Countries test tactics in 'war' against COVID-19. Science 2020 Mar 20;367(6484):1287-1288. [doi: 10.1126/science.367.6484.1287] [Medline: 32193299]
4. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
5. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, Imperial College COVID-19 Response Team. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. Nature 2020 Aug;584(7821):425-429. [doi: 10.1038/s41586-020-2488-1] [Medline: 32604404]
6. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
7. Di Bari M, Balzi D, Carreras G, Onder G. Extensive testing may reduce COVID-19 mortality: A lesson from northern Italy. Front Med (Lausanne) 2020;7:402 [FREE Full text] [doi: 10.3389/fmed.2020.00402] [Medline: 32766267]
8. Long C, Xu H, Shen Q, Zhang X, Fan B, Wang C, et al. Diagnosis of the coronavirus disease (COVID-19): rRT-PCR or CT? Eur J Radiol 2020 May;126:108961 [FREE Full text] [doi: 10.1016/j.ejrad.2020.108961] [Medline: 32229322]

9. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: Issues affecting the results. Expert Rev Mol Diagn 2020 May;20(5):453-454 [FREE Full text] [doi: 10.1080/14737159.2020.1757437] [Medline: 32297805]
10. Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Inform Med Unlocked 2020;20:100378 [FREE Full text] [doi: 10.1016/j.imu.2020.100378] [Medline: 32839734]
11. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
12. Pepe E, Bajardi P, Gauvin L, Privitera F, Lake B, Cattuto C, et al. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. Sci Data 2020 Jul 08;7(1):230 [FREE Full text] [doi: 10.1038/s41597-020-00575-2] [Medline: 32641758]
13. Rader B, Scarpino SV, Nande A, Hill AL, Adlam B, Reiner RC, et al. Crowding and the shape of COVID-19 epidemics. Nat Med 2020 Dec;26(12):1829-1834. [doi: 10.1038/s41591-020-1104-0] [Medline: 33020651]
14. Zens M, Brammertz A, Herpich J, Südkamp N, Hinterseer M. App-based tracking of self-reported COVID-19 symptoms: Analysis of questionnaire data. J Med Internet Res 2020 Sep 09;22(9):e21956 [FREE Full text] [doi: 10.2196/21956] [Medline: 32791493]
15. Corona-Warn-App. German Federal Government. URL: https://www.bundesregierung.de/breg-de/themen/corona-warn-app [accessed 2021-02-12]
16. Corona-Datenspende. Robert Koch Institut. URL: https://corona-datenspende.de [accessed 2021-02-12]
17. CovApp. https://covapp.charite.de. URL: https://covapp.charite.de [accessed 2021-02-12]
18. Warn-App NINA. Federal Office of Civil Protection and Disaster Assistance. URL: https://www.bbk.bund.de/DE/NINA/Warn-App_NINA_node.html [accessed 2021-02-12]
19. CoCoV. University Hospital Ulm. URL: https://sysbio.uni-ulm.de/?COVID-19_Software:CoCoV [accessed 2021-02-12]
20. Corona Check. University Hospital Würzburg. URL: https://www.coronacheck.science/en/ [accessed 2021-02-12]
21. Corona Health. University of Würzburg. URL: https://www.corona-health.net [accessed 2021-02-12]
22. Oberle D, Mentzer D, Weber G. Befragung zur Verträglichkeit der Impfstoffe gegen das neue Coronavirus (SARS-CoV-2) mittels Smartphone-App SafeVac 2. Paul-Ehrlich-Institut. 2020. URL: https://www.pei.de/SharedDocs/Downloads/DE/newsroom/bulletin-arzneimittelsicherheit/einzelartikel/2020-safevac-app.pdf?__blob=publicationFile&v=3 [accessed 2021-05-30]
23. Launch of the WHO Academy and the WHO Info mobile applications. World Health Organization. 2020 May 13. URL: https://www.who.int/news/item/13-05-2020-launch-of-the-who-academy-and-the-who-info-mobile-applications [accessed 2021-02-12]
24. Datenerfassung durch die Gutenberg COVID-19 App. University Medicine Mainz. 2020. URL: https://www.unimedizin-mainz.de/gcs/informationen-zur-studie/gutenberg-covid-19-app.html [accessed 2021-02-12]
25. CoronaBoXX. KAIROS. URL: https://www.kairos.de/en/products/coronaboxx/ [accessed 2021-05-28]
26. Mahase E. Covid-19: UK regulator approves lateral flow test for home use despite accuracy concerns. BMJ 2020 Dec 23;371:m4950. [doi: 10.1136/bmj.m4950] [Medline: 33361088]
27. Taylor-Phillips S, Berhane S, Sitch AJ, Freeman K, Price MJ, Davenport C, et al. Information given by websites selling home self-sampling COVID-19 tests: An analysis of accuracy and completeness. BMJ Open 2020 Nov 06;10(11):e042453 [FREE Full text] [doi: 10.1136/bmjopen-2020-042453] [Medline: 33158838]
28. Schmidt B, Sivanathan V, Schmiedel S, Kohnen W, Hoffmann D. The ambulatory management of COVID-19 via the German Department of Health. Dtsch Arztebl Int 2020 Jul 20;117(29-30):507-508 [FREE Full text] [doi: 10.3238/arztebl.2020.0507] [Medline: 33087231]
29. Kraus JM, Lausser L, Kuhn P, Jobst F, Bock M, Halanke C, et al. Big data and precision medicine: Challenges and strategies with healthcare data. Int J Data Sci Anal 2018 Jan 27;6(3):241-249. [doi: 10.1007/s41060-018-0095-0]
30. Völkel G, Wiese S, Holzmann K, Kraus JM, Schneider F, Görlach M, et al. TraqBio - Flexible progress tracking for core unit projects. PLoS One 2016;11(9):e0162857 [FREE Full text] [doi: 10.1371/journal.pone.0162857] [Medline: 27677174]
31. Halloway S, Bedra A. Programming Clojure. 2nd edition. Raleigh, NC: The Pragmatic Bookshelf; 2012.
32. Hipp RD. SQLite Release 3.29.0. SQLite. 2019. URL: https://sqlite.org/releaselog/3_29_0.html [accessed 2019-12-31]
33. Eastlake D, Crocker S, Schiller J. RFC 1750: Randomness Recommendations for Security. Marina del Rey, CA: RFC Editor, USC Information Sciences Institute; 1994 Dec. URL: https://dl.acm.org/doi/pdf/10.17487/RFC1750 [accessed 2021-05-28]
34. Cochran D. Twitter Bootstrap Web Development How-To. Birmingham, UK: Packt Publishing; 2012.
35. Spurlock J. Bootstrap. Sebastopol, CA: O'Reilly Media; 2013.
36. Duckett J. JavaScript and JQuery: Interactive Front-End Web Development. Hoboken, NJ: Wiley; 2014.
37. Fulmer J. Siege 4.0.4. JoeDog Software. 2017. URL: https://www.joedog.org/2017/09/11/siege-4-0-4/ [accessed 2021-05-30]
38. sysbio-bioinf. CTest. GitHub. 2020. URL: https://github.com/sysbio-bioinf/CTest

39. Höher M, Brummer T, Kestler H. Concept and initial experience with a mid-term DICOM archiving system acting as an intelligent buffer to clinical requests. In: Proceedings of Computers in Cardiology 1999.: IEEE; 1999 Presented at: Computers in Cardiology 1999; September 26-29, 1999; Hannover, Germany p. 319-322. [doi: 10.1109/cic.1999.825971]
40. Al Shamsi H, Almutairi AG, Al Mashrafi S, Al Kalbani T. Implications of language barriers for healthcare: A systematic review. Oman Med J 2020 Mar;35(2):e122 [FREE Full text] [doi: 10.5001/omj.2020.40] [Medline: 32411417]
41. Hunter-Adams J, Rother H. A qualitative study of language barriers between South African health care providers and cross-border migrants. BMC Health Serv Res 2017 Jan 31;17(1):97 [FREE Full text] [doi: 10.1186/s12913-017-2042-5] [Medline: 28143514]
42. Steinberg EM, Valenzuela-Araujo D, Zickafoose JS, Kieffer E, DeCamp LR. The "battle" of managing language barriers in health care. Clin Pediatr (Phila) 2016 Dec;55(14):1318-1327 [FREE Full text] [doi: 10.1177/0009922816629760] [Medline: 26896341]
43. Omar S, Bartz C, Becker S, Basenach S, Pfeifer S, Trapp C, Palatina Public Health Study Group. Duration of SARS-CoV-2 RNA detection in COVID-19 patients in home isolation, Rhineland-Palatinate, Germany, 2020 - An interval-censored survival analysis. Euro Surveill 2020 Jul;25(30):2001292 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.30.2001292] [Medline: 32734855]
44. Aleta A, Martín-Corral D, Pastore Y Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nat Hum Behav 2020 Sep;4(9):964-971. [doi: 10.1038/s41562-020-0931-9] [Medline: 32759985]
45. Rong X, Yang L, Chu H, Fan M. Effect of delay in diagnosis on transmission of COVID-19. Math Biosci Eng 2020 Mar 11;17(3):2725-2740 [FREE Full text] [doi: 10.3934/mbe.2020149] [Medline: 32233563]
46. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. N Engl J Med 2020 Mar 19;382(12):1177-1179. [doi: 10.1056/nejmc2001737]
47. Miles A. The psychological implications of diagnostic delay in colorectal cancer patients. In: Olsson L, editor. Timely Diagnosis of Colorectal Cancer. Cham, Switzerland: Springer; 2018:103-119.

Abbreviations

BMBF: Federal Ministry of Education and Research
CSV: comma-separated values
DFG: German Science Foundation
HTTPS: HTTP Secure
IP: Internet Protocol
JSON: JavaScript Object Notation
NFN: Nationales Forschungs Netzwerk
QR: Quick Response
RT-PCR: reverse transcription–polymerase chain reaction
SFTP: Secure File Transfer Protocol
SSL: Secure Sockets Layer
ZIK: Zentrum für Information und Kommunikation
ZIV: Zentrum für Innovative Versorgung
ZPM: Zentren für Personalisierte Medizin

Edited by C Basch; submitted 22.01.21; peer-reviewed by R Pryss, M Boeker; comments to author 08.02.21; revised version received 23.02.21; accepted 12.05.21; published 07.06.21.

Please cite as:
Völkel G, Fürstberger A, Schwab JD, Werle SD, Ikonomi N, Gscheidmeier T, Kraus JM, Groß A, Holderried M, Balig J, Jobst F, Kuhn P, Kuhn KA, Kohlbacher O, Kaisers UX, Seufferlein T, Kestler HA
Patient Empowerment During the COVID-19 Pandemic by Ensuring Safe and Fast Communication of Test Results: Implementation and Performance of a Tracking System
J Med Internet Res 2021;23(6):e27348
URL: https://www.jmir.org/2021/6/e27348
doi:10.2196/27348
PMID:33999836

Original Paper

Using Fear and Anxiety Related to COVID-19 to Predict Cyberchondria: Cross-sectional Survey Study

Xue Wu[1], PhD; Nabi Nazari[2], PhD; Mark D Griffiths[3], PhD

[1]School of Economics and Management, Kaili University, Guizhou, China

[2]Department of Psychology, Faculty of Literature and Human Sciences, Lorestan University, Khorramabad, Iran

[3]International Gaming Research Unit, Psychology Division, Nottingham Trent University, Nottingham, United Kingdom

Corresponding Author:
Nabi Nazari, PhD
Department of Psychology
Faculty of Literature and Human Sciences
Lorestan University
PO Box 465
Khorramabad
Iran
Phone: 98 09197119027
Fax: 98 6123986253
Email: Nazariirani@gmail.com

Abstract

Background: Studies have highlighted that fear and anxiety generated by COVID-19 are important psychological factors that affect all populations. There currently remains a lack of research on specific amplification factors regarding fear and anxiety in the context of the COVID-19 pandemic. Despite established associations between anxiety sensitivity, intolerance of uncertainty, and cyberchondria, empirical data investigating the associations between these three variables, particularly in the context of the COVID-19 pandemic, are currently lacking. Urgent research is needed to better understand the role of repeated media consumption concerning COVID-19 in amplifying fear and anxiety related to COVID-19.

Objective: This study investigated the associations between fear of COVID-19, COVID-19 anxiety, and cyberchondria.

Methods: Convenience sampling was used to recruit respondents to participate in an online survey. The survey, which was distributed via social media and academic forums, comprised the Cyberchondria Severity Scale, Fear of COVID-19 Scale, Coronavirus Anxiety Scale, Anxiety Sensitivity Index, and Intolerance of Uncertainty Scale. Multiple mediation analyses were conducted using structural equation modeling.

Results: A total of 694 respondents (males: n=343, females: n=351) completed the online survey. The results showed that fear and anxiety generated by COVID-19 predicted cyberchondria (fear: β=.39, SE 0.04, *P*<.001, *t*=11.16, 95% CI 0.31-0.45; anxiety: β=.25, SE 0.03, *P*<.001, *t*=7.67, 95% CI 0.19-0.32). In addition, intolerance of uncertainty and anxiety sensitivity mediated the relationship between fear and anxiety generated by COVID-19 with cyberchondria. In a reciprocal model, the standardized total effects of cyberchondria on fear of COVID-19 (β=.45, SE 0.04, *P*<.001, *t*=15.31, 95% CI 0.39-0.51) and COVID-19 anxiety (β=.36, SE 0.03, *P*<.001, *t*=11.29, 95% CI 0.30-0.41) were statistically significant, with moderate effect sizes. Compared to males, females obtained significantly higher scores for cyberchondria ($t_{1,692}$=–2.85, *P*=.004, Cohen *d*=0.22), COVID-19 anxiety ($t_{1,692}$=–3.32, *P*<.001, Cohen *d*=0.26), and anxiety sensitivity ($t_{1,692}$=–3.69, *P*<.001, Cohen *d*=0.29).

Conclusions: The findings provide a better understanding of the role of COVID-19 in amplifying cyberchondria. Based on these results, cyberchondria must be viewed as a significant public health issue. Importantly, increasing awareness about cyberchondria and online behavior at both the individual and collective levels must be prioritized to enhance preparedness and to reduce the adverse effects of current and future medical crises.

(J Med Internet Res 2021;23(6):e26285) doi:10.2196/26285

KEYWORDS

COVID-19; cyberchondria; COVID-19 fear; COVID-19 anxiety; anxiety sensitivity; anxiety; intolerance of uncertainty; mental health; survey; SEM

Introduction

Background

The COVID-19 outbreak is more of a global emergency than a medical challenge. Research highlights the intense and broad spectrum of psychosocial ramifications that pandemics can inflict on the general population [1]. Fear of COVID-19 and COVID-19 anxiety, coupled with quarantine and isolation [2], can generate specific negative psychological responses such as maladaptive behaviors, emotional distress, and avoidance reactions among both general and patient populations [3,4]. Compared with previous pandemics (eg, severe acute respiratory syndrome [SARS]), psychological distress and anxiety disorders related to the increasingly widespread use of the internet are relatively novel problems in psychiatric and medical settings. The internet can be a useful source of health information [5] and has become increasingly prevalent among all members of the public. However, repeated media exposure to pandemic-related information and excessive searching for health-related information on the internet can significantly exacerbate anxiety and create an escalating pattern of psychological distress that is difficult to break. This has been termed “cyberchondria” [6,7].

Cyberchondria in the Context of the COVID-19 Pandemic

Cyberchondria has been defined as “anxiety resulting from a health-related search online” [8,9]. Cyberchondria is conceptualized as a multidimensional construct, including repetitive (excessive) online searching for health-related information, distress (increased negative affect), compulsion (interruption of daily routine), and reassurance seeking. Seeking health-related information on the internet to reduce anxiety may result in more anxiety or distress [10]. In the latest (fifth) edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), individuals can receive one of two anxiety-related diagnoses. Somatic symptom disorder refers to anxiety concerning health diagnoses in the context of significant somatic symptoms whereas illness anxiety disorder refers to a health anxiety diagnosis without somatic symptoms. Individuals affected by these mental health problems tend to misinterpret minor bodily sensations or symptoms as if they were severe illnesses [11]. When individuals with these diagnoses use online networks to search for health-related information, they are classified as having cyberchondria [12]. As a safety behavior, cyberchondria (ie, anxiety when searching online for health-related information) may fuel psychopathological vulnerabilities [12]. This indicates a strong relationship between health anxiety and cyberchondria [13,14]. Individuals with elevated health anxiety exhibit higher anxiety levels during and after online health-related searches and report more frequent and longer online searches than those with lower levels of health anxiety [15,16]. These discrepancies between the purpose and outcome of cyberchondria may complicate our understanding of patterns of cyberchondria. In addition to anxiety, fear is a substantial motivating factor in seeking health-related information [17]. Cyberchondria may also exacerbate fear of illness and have a negative impact on relationships with primary care physicians. Cyberchondria is also associated with accessing increased health resources, as measured by the number of visits to general practitioners and other health professionals [18].

Fear in the Context of the COVID-19 Pandemic

Fear, generated by trauma-related stimuli, is a prominent emotion in psychopathology [19]. Fear can be viewed as an adaptive response to threat, which can be a motivating factor that facilitates protective and preventive behavior among individuals to avoid infection and follow pandemic-related health instructions. However, for individuals who experience fear intensely, it may result in an elevated risk perception. Consequently, this adaptive response to fear becomes maladaptive when these emotional responses fail to provide accurate information. Prior experience, cognitive and attentional biases, and mental disorders can all generate faulty appraisals of the physical and social environment, leading to maladaptive emotional reactions. In the context of the current COVID-19 pandemic, fear and anxiety can elicit additional media consumption [20]. In addition, the consumption of pandemic-related media coverage may be an important factor that is associated with anxiety and psychological distress [7,21]. For example, exposure to warning messages, as well as inaccurate and misleading information concerning life-threatening aspects of COVID-19 during online searching, can exacerbate anxiety and worries related to the pandemic [22]. Moreover, a significant positive association has been found between anxiety resulting from online health searches for oneself and anxiety resulting from online health searches for others [8]. Fear of self-infection or infecting family members is one of the most common reactions to pandemics [23,24] and can result in health anxiety, worries, specific phobias, and psychological distress [25-27]. There is also a strong relationship between cyberchondria and health anxiety [28]. Health anxiety can motivate excessive or repeated health-related information seeking on the internet, which can amplify anxiety or distress (eg, fear) [9,15,29]. Therefore, individuals with cyberchondria may be anxious about the health of family members, attempt to diagnose them online, and/or take additional measures as a consequence of their fear of COVID-19.

Uncertainty and Anxiety Sensitivity During the COVID-19 Pandemic

The current pandemic has caused much uncertainty about many different aspects of daily life. Intolerance of uncertainty is recognized as a strong predictor of cyberchondria. In times of uncertainty, reducing uncertainty has a central role in motivating searching for health information on the internet [30]. In addition, individuals with a higher level of intolerance of uncertainty exhibit prospective anxiety due to dispositional fear of unknown future events [31,32]. Individuals with a higher intolerance of uncertainty levels may perceive uncertain situations as both threatening and aversive. Therefore, individuals engage in uncertainty-reducing behaviors (eg, repeatedly seeking reassurance due to worries) to moderate the perceptions of uncertainty and threat [33].

Anxiety sensitivity, defined as “the fear of sensations of anxious arousal based on beliefs about their harmful consequences,” is conceptualized as a cognitive-emotional individual difference factor of the fear related to bodily sensations [34]. Experimental studies have demonstrated a positive association between anxiety sensitivity and oversearching of medical information. Compared to individuals with generalized anxiety disorder, anxiety sensitivity is recognized as a potential risk factor for increased anxiety related to COVID-19 [35]. Research also indicates that accurate knowledge about pandemics may be associated with anxiety. During the COVID-19 pandemic, it is possible that such individuals search for information and medical news related to COVID-19 with high sensitivity to anxiety, which leads to an increase in their fear.

Purpose of This Study

Understanding COVID-19 pandemic–related psychopathology development is limited due to numerous individual and contextual factors. There is currently a lack of research on specific amplification factors regarding fear and anxiety in the context of the COVID-19 pandemic. Cyberchondria is a maladaptive behavioral pattern, more likely during public health crises such as the COVID-19 pandemic. Research is urgently needed to better understand the role of repeated media consumption concerning COVID-19 [36]. Despite established associations between anxiety sensitivity, intolerance of uncertainty, and online medical information seeking [37-39], there is currently a lack of empirical data concerning the associations between anxiety sensitivity, intolerance of uncertainty, and cyberchondria, particularly in the context of the COVID-19 pandemic.

Therefore, this study investigated the associations between fear of COVID-19, COVID-19 anxiety, and cyberchondria. Using structural equation modeling (SEM), a mediation analysis was carried out to investigate the underlying mechanism between cyberchondria and fear of COVID-19, COVID-19 anxiety, intolerance of uncertainty, and anxiety sensitivity. In addition, an evaluation of the reverse mediation model of the association between the study variables was also carried out. The study may potentially contribute to a better understanding of the pandemic in relation to cyberchondria. Moreover, the findings provide additional insight into cyberchondria and the pandemic, providing important information to clinical practitioners and policymakers. It was hypothesized that intolerance of uncertainty and anxiety sensitivity would mediate the association between fear of COVID-19 and COVID-19 anxiety on cyberchondria.

Methods

Inclusion Criteria

The eligibility criteria included (1) age >18 years, (2) not hospitalized or quarantined in the current or a past viral pandemic due to infection, (3) not having (or suspect as having) COVID-19, (4) being able to read and complete an online survey and provide informed consent, (5) fluency in the Persian language, and (6) currently living in Iran. Only completed questionnaires were analyzed.

Measures

The survey comprised the Cyberchondria Severity Scale, Fear of COVID-19 Scale, Coronavirus Anxiety Scale, Anxiety Sensitivity Index, and Intolerance of Uncertainty Scale.

Cyberchondria Severity Scale–Short Form

The Cyberchondria Severity Scale–Short Form (CSS-12) [40] is a 12-item self-report scale designed to assess anxiety attributable to health-related online searches. The items (eg, “If I notice an unexplained bodily sensation, I will search for it on the Internet”) are rated on 5-point scale from 1 (never) to 5 (always). The scale comprises three subscales: compulsion, distress, and mistrust of medical professionals. Higher scores indicate greater cyberchondria. The internal consistency of the CSS-12 in this study was excellent (Cronbach α=.90).

Fear of COVID-19 Scale

The Fear of COVID-19 Scale (FCV-19S) [41] is a 7-item unidimensional scale that assesses fear of COVID-19. The items (eg, “I am afraid of losing my life because of COVID-19”) are rated on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). Higher scores indicate a greater fear of COVID-19. The internal consistency of the FCV-19S in this study was very good (Cronbach α=.83).

Coronavirus Anxiety Scale

The 5-item Coronavirus Anxiety Scale (CAS) [42] assesses dysfunctional anxiety associated with COVID-19. The items (eg, “I had trouble falling or staying asleep because I was thinking about the coronavirus”) are rated on a 5-point scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Higher scores are associated with a COVID-19 anxiety diagnosis, impairment, maladaptive coping, and suicidal ideation. The internal consistency of the CAS in this study was excellent (Cronbach α=.90).

The Anxiety Sensitivity Index-3

The Anxiety Sensitivity Index-3 (ASI-3) [43] is an 18-item self-report scale that assesses anxiety-related symptoms. Items (eg, “It scares me when my heart beats rapidly”) are rated on a 5-point scale from 0 (not at all) to 4 (very much). Higher scores indicate a more severe anxiety sensitivity level. The internal consistency of the ASI-3 in this study was very good (Cronbach α=.85).

Intolerance of Uncertainty Scale-12

The Intolerance of Uncertainty Scale-12 (IUS-12) [44] is a 12-item scale that assesses individuals’ responses to uncertainty. The items (eg, “It frustrates me not having all the information I need”) are rated on a 5-point scale from 1 (not at all characteristic of me) to 5 (entirely characteristic of me). Higher scores indicate greater uncertainty. The internal consistency of the IUS-12 in this study was very good (Cronbach α=.82).

Participant Recruitment

The study was conducted during the COVID-19 pandemic (October and November 2020) via convenience sampling; hence, all data were collected online because face-to-face data collection was not possible. The participants were recruited over a 6-week period using an online platform to complete the survey. The recruitment process included advertising the study via social media platforms (Instagram, WhatsApp) with links to the survey. In addition, the link was distributed on several academic forums. Once the link was clicked, it led to an informed consent page to be read and agreed upon before proceeding to the survey. Only those who provided informed consent were able to access the survey. The informed consent page included information about the study goals, such as the study's objectives and confidentiality.

Sample Size

A priori power analysis for multiple linear regression was calculated using G*Power (Heinrich-Heine-Universität Düsseldorf) to determine the sample size, with an alpha of .025, a power of 0.80, Cohen f^2 of 0.02, and two predictors [45]. The effect size value (Cohen f^2=0.02) signifies small effect sizes, according to Cohen's guidelines [46]. The desired total sample size was 576. In total, 694 participants were recruited in this study, which allowed for a 20% data attrition.

Ethics

The study, including all assessments and procedures for the study, were reviewed by the National Institute for Medical Research and Development and the Institutional Human Research Ethics Committee. The corresponding author's institutional review board also approved the research protocol to ensure participant confidentiality, sampling, and informed consent.

Data Analysis

Descriptive Statistics

Descriptive statistics were used to calculate the sample characteristics. Absolute skewness and kurtosis values assessed the normality assumption [47]. Variance inflation factor (VIF) was utilized to examine multicollinearity (1<VIF<3) [48]. Pearson coefficient correlation analysis was carried out to calculate the association between cyberchondria and the study variables. There were no missing values in the assessed variables. Therefore, no imputation method was implemented.

Multiple Mediation Analysis

Parallel multiple mediation analysis was conducted using SEM with a 95% CI for indirect effects and 5000 bootstrapping [49]. Once the measurement models were fitted to the data, 2 SEM models were examined. The first SEM investigated the relationships between fear of COVID-19 and COVID-19 anxiety (as independent variables) with cyberchondria. Anxiety sensitivity and intolerance of uncertainty were potential mediators. In the second SEM, the reverse model was examined. An indirect effect was considered statistically significant when the bias-corrected CI does not include zero [50,51]. Cohen f^2 values of ≥0.15 and ≥0.35 signified approximately moderate to large effect sizes, according to Cohen's guidelines [46]. SPSS (version 25, IBM Corp) and AMOS (version 24, IBM Corp) were utilized to test hypothesizes (two-tailed), and an alpha level of .05 was used to indicate statistical significance.

Results

Descriptive Statistics

Of the 820 returned surveys, 694 met the inclusion criteria and were included in the analysis. Therefore, the sample comprised 694 adults from the general population (males: n=343, 49.4%; females: n=351, 50.6%) , with a mean age of 27.92 years (SD 5.22, range 19-41 years). The demographic characteristics of the sample are shown in Table 1.

There was no significant difference between males and females (χ^2=0.09, *P*=.76). The participants were well educated and young. With respect to educational level, 25.8% (n=179) had completed high school, 49.2% (n=341) had a bachelor's degree, and 25% (n=174) had a master's and/or higher degree. Compared to males, females had significantly higher scores for cyberchondria ($t_{1,692}$=–2.85, *P*=.004, Cohen *d*=0.22), COVID-19 anxiety ($t_{1,692}$=–3.32, *P*<.001, Cohen *d*=0.26), and anxiety sensitivity ($t_{1,692}$=–3.69, *P*<.001, Cohen *d*=0.29). Compared to females, males had significantly higher scores for intolerance of uncertainty ($t_{1,692}$=2.29, *P*=.02, Cohen *d*=0.18). Gender differences for other variables were nonsignificant (*P*>.05) (Table 1).

The univariate normality of the data was checked. Values of skewness and kurtosis were within <|1|, suggesting the absence of severe normality. The VIF values demonstrated no violation of multicollinearity (Table 2). The Pearson coefficient correlation analyses showed a moderate to large correlations between variables (Table 2). Correlation analysis revealed a moderate to large correlation between cyberchondria and fear of COVID-19, COVID-19 anxiety, anxiety sensitivity, and intolerance of uncertainty.

Table 1. Demographic characteristics and descriptive statistics of the sample (N=694).

Characteristic	Participants	Statistics[a]	*P* value
Gender, n (%)		χ^2=0.09	.76
Male	343 (49.4)		
Female	351 (50.6)		
Age group, n (%)		χ^2=11.60	.003
19-25 years	193 (27.8)		
26-31 years	266 (38.3)		
>31 years	235 (33.9)		
Age (years), mean (SD)	27.92 (5.29)	$t_{1,692}$=0.82	.41
Cyberchondria, mean (SD)	33.17 (8.04)	$t_{1,692}$=–2.85	004
Coronavirus anxiety, mean (SD)	11.91 (2.51)	$t_{1,692}$=–3.32	.001
Fear of COVID-19, mean (SD)	15.21 (4.93)	$t_{1,692}$=–1.70	.09
Intolerance of uncertainty, mean (SD)	38.82 (9.52)	$t_{1,692}$=2.29	.02
Anxiety sensitivity, mean (SD)	28.52 (8.29)	$t_{1,692}$= –3.69	<.001

[a] Negative *t* values indicate that females obtained higher scores.

Table 2. Correlation matrix of main variables (N=694).

Item	1	2	3	4	5	6	Skewness	Kurtosis	VIF[a]
1. Age	1.00						0.49	–0.99	—[b]
2. Cyberchondria	–0.09[c]	1.00					0.38	0.85	2.51
3. Coronavirus anxiety	0.04	0.36[d]	1.00				0.70	–0.69	1.82
4. Fear of COVID-19	–0.06	0.46[d]	0.29[d]	1.00			0.39	0.75	1.35
5. Intolerance of uncertainty	0.26[d]	0.43[d]	0.31[d]	0.29[d]	1.00		0.65	–0.96	2.23
6. Anxiety sensitivity	–0.03	0.31[d]	0.27[d]	0.44[d]	0.36[d]	1.00	0.22	–1.02	1.92

[a] VIF: variance inflation factor.

[b] Not applicable.

[c] Correlation significant at the *P*<.05 level (two-tailed).

[d] Correlation significant at the *P*<.01 level (two-tailed).

Multiple Mediation Analysis

The first SEM mediation analysis showed that fear of COVID-19 and COVID-19 anxiety were significantly associated with cyberchondria via both direct and indirect paths. The standardized total effects for fear of COVID-19 and COVID-19 anxiety on cyberchondria were both statistically significant (fear: β=.39, SE 0.04, *P*<.001, *t*=11.16, 95% CI 0.31-0.45; anxiety: β=.25, SE 0.03, *P*<.001, *t*=7.67, 95% CI 0.19-0.32). Anxiety sensitivity and intolerance of uncertainty were also positively associated with cyberchondria (Figure 1). In addition, for fear of COVID-19 and COVID-19 anxiety, the indirect effects of anxiety sensitivity and intolerance of uncertainty on cyberchondria were significant (Table 3).

In the reciprocal model, the standardized total effects of cyberchondria on both fear of COVID-19 and COVID-19 anxiety were statistically significant (fear: β=.45, SE 0.04, *P*<.001, *t*=15.31, 95% CI 0.39-0.51; anxiety: β=.36, SE 0.03, *P*<.001, *t*=11.29, 95% CI 0.30-0.41. The indirect effects of cyberchondria on COVID-19 anxiety via anxiety sensitivity and intolerance of uncertainty were both significant. However, only cyberchondria's indirect effect on fear of COVID-19 via anxiety sensitivity was significant (Figure 2). The effect sizes for associations in both mediation models are reported in Table 4.

Figure 1. The multiple mediation model. Path coefficient: standardized coefficient (*t* value). ****P*<.001 level (two-tailed).

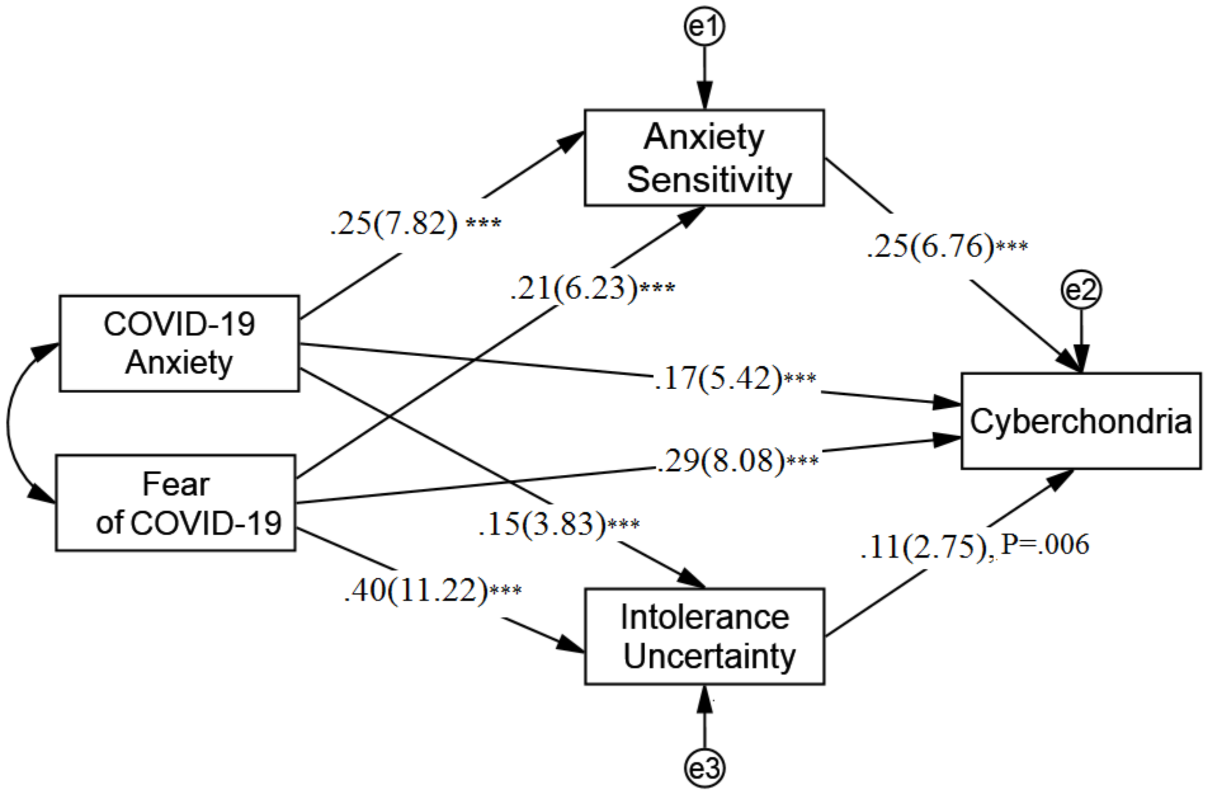


Table 3. Standardized indirect effects of model paths.

Model and indirect path[a]	Beta[b]	SE	*t* value	*P* value	95% CI
Model 1					
Coronavirus anxiety –> intolerance of uncertainty –> cyberchondria	.06	0.02	4.98	<.001	0.04 to 0.09
Fear of COVID-19 –> intolerance of uncertainty –> cyberchondria	.06	0.01	4.52	<.001	0.03 to 0.08
Coronavirus anxiety –> anxiety sensitivity –> cyberchondria	.02	0.01	2.23	.03	0.004 to 0.03
Fear of COVID-19 –> anxiety sensitivity –> cyberchondria	.04	0.02	2.67	.01	0.01 to 0.07
Model 2 (reciprocal model)					
Cyberchondria –> intolerance of uncertainty –> coronavirus anxiety	.07	0.02	4.26	<.001	0.04 to 0.11
Cyberchondria –> anxiety sensitivity –> coronavirus anxiety	.04	0.01	2.98	.004	0.01 to 0.07
Cyberchondria –> intolerance of uncertainty –> fear of COVID-19	.02	0.02	1.15	.25	–0.01 to 0.05
Cyberchondria –> anxiety sensitivity –> fear of COVID-19	.11	0.02	6.99	<.001	0.08 to 0.15

[a] Arrow indicates path direction.

[b] Beta: standardized path coefficient.

Figure 2. The reciprocal multiple mediation model. Path coefficient: standardized coefficient (*t* value), dash line: nonsignificant path. ***$P<.001$ level (two-tailed).

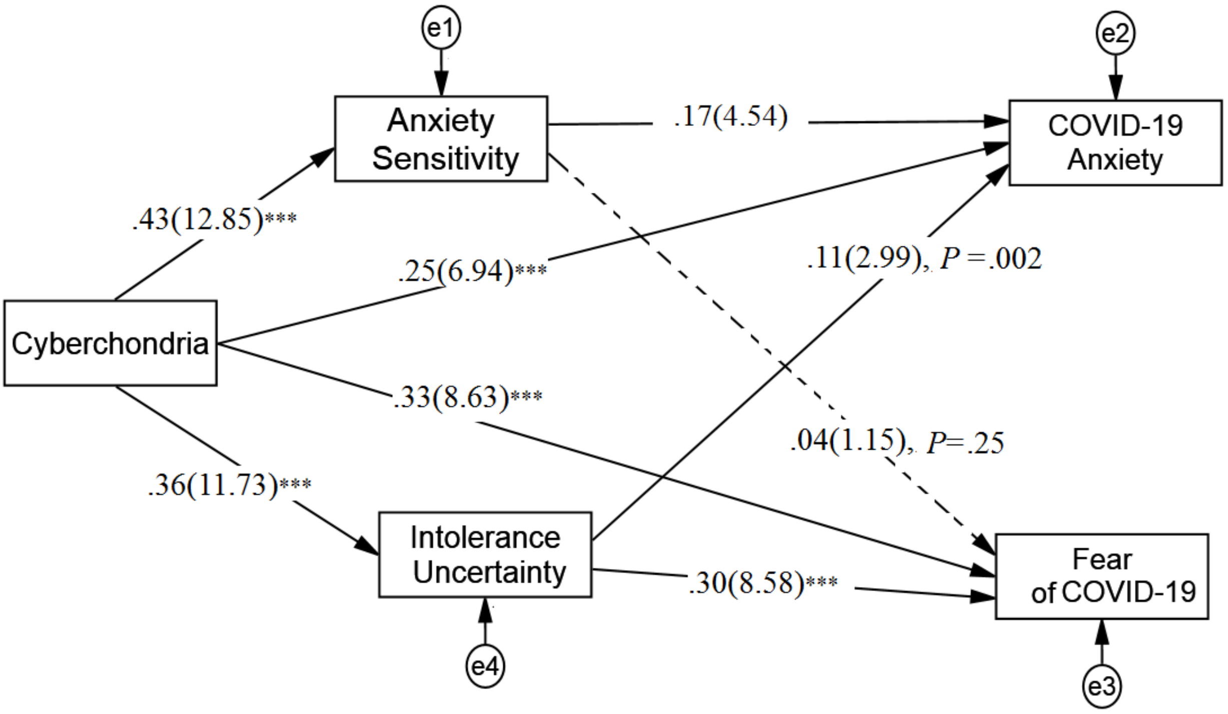


Table 4. Effect sizes of model paths.

Model and path[a]	Cohen f^2	*P* value	95% CI
Model 1			
Fear of COVID-19 –> cyberchondria	0.10	<.001	0.06-0.16
Fear of COVID-19 –> intolerance of uncertainty	0.06	.002	0.03-0.09
Fear of COVID-19 –> anxiety sensitivity	0.19	<.001	0.13-0.27
Coronavirus anxiety –> cyberchondria	0.04	.008	0.02-0.07
Coronavirus anxiety –> intolerance of uncertainty	0.07	.001	0.04-0.11
Coronavirus anxiety –> anxiety sensitivity	0.03	.04	0.01-0.06
Intolerance of uncertainty –> cyberchondria	0.08	.002	0.04-0.14
Anxiety sensitivity –> cyberchondria	0.01	.18	0.001-0.04
Model 2 (reciprocal model)			
Cyberchondria –> coronavirus anxiety	0.06	.001	0.03-0.10
Cyberchondria –> fear of COVID-19	0.12	<.001	0.07-0.19
Cyberchondria –> intolerance of uncertainty	0.23	<.001	0.15-0.31
Cyberchondria –> anxiety sensitivity	0.16	<.001	0.11-0.25
Intolerance of uncertainty –> coronavirus anxiety	0.02	.03	0.01-0.05
Intolerance of uncertainty –> fear of COVID-19	0.002	.65	0.00-0.01
Anxiety sensitivity –> coronavirus anxiety	0.01	.15	0.001-0.03
Anxiety sensitivity –> fear of COVID-19	0.11	<.001	0.06-0.17

[a] Arrow indicates path direction.

Discussion

Principal Findings

Understanding COVID-19 pandemic–related psychopathological development has been limited due to numerous individual and contextual factors. There is concern that individuals affected by fear and anxiety generated by the pandemic will rapidly outnumber infected cases. This study investigated the impact of the relationship between cyberchondria and the fear and anxiety generated by COVID-19. The study also explored whether the relationships would be mediated by the intolerance of uncertainty and anxiety sensitivity. Results indicated that greater fear and anxiety related to COVID-19 directly predicted cyberchondria [52-54] and indirectly via the mediator variables. Cyberchondria and anxiety generated by COVID-19 was bidirectional. A higher level of cyberchondria directly predicted a higher level of fear and anxiety generated by COVID-19, and indirectly via the mediators.

The mediation analysis empirically showed that intolerance of uncertainty and anxiety sensitivity mediated the associations between fear and anxiety generated by COVID-19 on cyberchondria. In the context of COVID-19, individuals with a higher intolerance of uncertainty levels may search for medical information to reduce uncertainty that results in additional negative experience [55-57]. Intolerance of uncertainty amplifies both threat perception and uncertainty perception [58], which can lead to more engagement in safety behaviors (eg, checking behavior) [59]. Like a vicious circle, seeking health-related information on the internet to reduce uncertainty may be associated with greater levels of uncertainty and therefore amplify health anxiety. These safety-seeking behaviors to reduce uncertainty are often not long lasting. In addition, most uncertainty reduction attempts meet with information that is overly brief and inaccurate, which can lead to greater levels of uncertainty. Moreover, uncertainty-reducing behaviors can lead to greater uncertainty perceptions and/or higher perceived threat severity. During a pandemic, individuals with a high intolerance of uncertainty perceive low-risk situations as highly threatening and report higher anxiety levels [34].

Individuals with high anxiety sensitivity (1) are more afraid of pain and more likely to seek unnecessary treatment for minor pain symptoms [60]; and (2) may misinterpret symptoms, which can result in bodily sensations related to anxiety (eg, sweating, shaky hands) that may be interpreted as severe physical symptoms or illness [61]. Since individuals with elevated anxiety sensitivity interpret anxiety-related bodily sensations to be dangerous, they exhibit increased online medical information-seeking behaviors in an attempt to placate concerns about the origins of anxiety-related bodily sensations. Specifically, this cognitive-affective condition is considered conceptually distinct from anxiety and reflects the fear of anxious arousal symptoms. Anxiety sensitivity is a more robust predictor of posttraumatic symptoms with a moderate to large effect size [62]. On the other hand, the reduction in anxiety sensitivity positively predicts a reduction in the severity of anxiety symptoms [63,64]. This theoretical process is consistent with research that has found the engagement in safety behaviors, including using medical websites to investigate medical symptoms, results in increased levels of health anxiety. With respect to the bidirectional associations between intolerance of uncertainty and anxiety sensitivity, findings suggest that evaluation of both conditions can add to the diagnosis process for individuals with cyberchondria.

This study also found some gender differences. In line with the few studies to date, females reported a higher levels for cyberchondria than males [65]. Females also reported higher levels of COVID-19 anxiety and anxiety sensitivity than males in this study. In line with recent studies, females reported more psychological problems associated with COVID-19 than males [66-68]. Our findings concur with previous studies indicating that females report greater psychological problems and are more likely to develop anxiety symptoms than males [69]. Regarding age, the results suggest that behaviors related to cyberchondria appear to be more prevalent among younger individuals [70,71].

Limitations

The findings of this study should be interpreted in light of several limitations. This study was conducted during October 2020 at the height of the COVID-19 pandemic in Iran. Therefore, to minimize infection risk, online data collection was utilized rather than a traditional face-to-face method. Online data collection may limit the participation of specific relevant population groups (eg, disadvantaged groups such as those living in poverty who may not have internet access). Therefore, the data do not represent all groups' views, affecting the generalizability of the study's findings. However, online data collection tends to provide more honest and truthful responses than offline methods [72]. Moreover, all data were self-reported and are therefore subject to well-established method biases. It should also be noted that the data collected did not include some potentially important variables such as whether (1) the participants were currently working or whether they had lost their job as a result of the pandemic, (2) they and/or their family members had experienced COVID-19, and (3) whether they had financial problems as a result of the pandemic. These are all variables that could be considered in future research when reexamining the variables of this study. Finally, the data were cross-sectional; therefore, determining the true relationships and directions of causality between the study variables is not possible. Future studies would need longitudinal designs to determine true causality.

Conclusion

Despite these limitations, the findings of the first SEM suggest that anxiety sensitivity and/or intolerance of uncertainty may lead to the development of cyberchondria in the context of the COVID-19 pandemic, with small to moderate effect sizes. In this study, greater fear and anxiety of COVID-19 were associated with greater cyberchondria. However, the reverse SEM demonstrated that cyberchondria is also associated with the study construct, with moderate to large effect sizes. In addition to the pandemic, anxiety sensitivity and intolerance of uncertainty can be critical in increasing or maintaining psychopathological development, as well as physical and psychological dysfunctions [73-75]. Given the co-occurring nature of mental health problems during the current pandemic, disorder-specific interventions may be difficult to justify when the clinical reality is complex and comorbidities are the norm [76]. Many clinical studies have recommended shifting from the traditional specific disorder–focused approach toward a transdiagnostic treatment as an alternative approach. Despite some scholars' assertions [77], the findings of this study do not justify cyberchondria as a transdiagnostic condition. However, the transdiagnostic treatments or application of the relative modules may be effective in treatment for cyberchondria. For example, higher self-awareness and contextual awareness enable individuals to clearly identify the triggering of negative responses and can help reduce maladaptive cognitive patterns by facilitating awareness or attention toward an object (eg, heartbeat, breathing) in a mindful manner [78,79]. To provide adaptive emotional responding to anxiety-related bodily sensations, self-awareness can be promoted by the mindfulness-based stress reduction therapy or other unified protocols [80].

This study's findings help to explain how the consequences of the pandemic can be associated with cyberchondria. The role of cyberchondria in the exacerbation of pandemic-related psychological distress can provide further evidence that maladaptive new age issues related to human-internet interaction need further attention from scholars, policymakers, and health care practitioners. Finally, based on the findings here, cyberchondria must be viewed as a significant public health issue. Importantly, increasing awareness about cyberchondria and online behavior at both the individual and collective levels must be prioritized to enhance preparedness and reduce adverse effects associated with the current pandemic and future medical crises.

Acknowledgments

This research was supported by the following special projects: "Response, Management and Impact of Public Health Events Such as the COVID-19 Epidemic," Kaili University (2020)—Empirical Investigation and Research on the Impact of the COVID-19 Epidemic on Key Industries in Guizhou (project number: YQZX201902); and "Highlight the Characteristics of Running a School," Kaili University (2020) (engineering subjects)—A Study on the Value Perception and Purchase Behavior of Camellia Oil in Southeast Guizhou: Based on Consumer Surveys (project number: 2020gkzs02).

Conflicts of Interest

None declared.

References

1. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate Psychological Responses and Associated Factors during the Initial Stage of the 2019 Coronavirus Disease (COVID-19) Epidemic among the General Population in China. Int J Environ Res Public Health 2020 Mar 06;17(5) [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
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 14;395(10227):912-920. [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
3. Asmundson GJ, Taylor S. Coronaphobia revisted: A state-of-the-art on pandemic-related fear, anxiety, and stress. J Anxiety Disord 2020 Dec;76:102326 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102326] [Medline: 33142124]
4. Nazari N, Zekiy AO, Feng L, Griffiths MD. Psychometric Validation of the Persian Version of the COVID-19-Related Psychological Distress Scale and Association with COVID-19 Fear, COVID-19 Anxiety, Optimism, and Lack of Resilience. Int J Ment Health Addict 2021 May 14:1-16 [FREE Full text] [doi: 10.1007/s11469-021-00540-z] [Medline: 34007255]
5. Riehm KE, Holingue C, Kalb LG, Bennett D, Kapteyn A, Jiang Q, et al. Associations Between Media Exposure and Mental Distress Among U.S. Adults at the Beginning of the COVID-19 Pandemic. Am J Prev Med 2020 Nov;59(5):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2020.06.008] [Medline: 33011008]
6. Elhai JD, Yang H, McKay D, Asmundson GJ. COVID-19 anxiety symptoms associated with problematic smartphone use severity in Chinese adults. J Affect Disord 2020 Sep 01;274:576-582 [FREE Full text] [doi: 10.1016/j.jad.2020.05.080] [Medline: 32663990]
7. Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357 [FREE Full text] [doi: 10.1037/hea0000875] [Medline: 32202824]
8. Starcevic V, Berle D. Cyberchondria: An old phenomenon in a new guise? In: Aboujaoude E, Starcevic V, editors. Mental Health in the Digital Age: Grave Dangers, Great Promise. Oxford, UK: Oxford University Press; 2015.
9. Starcevic V, Berle D. Cyberchondria: towards a better understanding of excessive health-related Internet use. Expert Rev Neurother 2013 Feb;13(2):205-213. [doi: 10.1586/ern.12.162] [Medline: 23368807]
10. Olatunji BO, Kauffman BY, Meltzer S, Davis ML, Smits JAJ, Powers MB. Cognitive-behavioral therapy for hypochondriasis/health anxiety: a meta-analysis of treatment outcome and moderators. Behav Res Ther 2014 Jul;58:65-74. [doi: 10.1016/j.brat.2014.05.002] [Medline: 24954212]
11. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. 5th ed. Virginia, US: American Psychiatric Association; 2013.

12. Vismara M, Caricasole V, Starcevic V, Cinosi E, Dell'Osso B, Martinotti G, et al. Is cyberchondria a new transdiagnostic digital compulsive syndrome? A systematic review of the evidence. Compr Psychiatry 2020 May;99:152167 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152167] [Medline: 32146315]
13. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. J Affect Disord 2019 Dec 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
14. Nekliudov NA, Blyuss O, Cheung KY, Petrou L, Genuneit J, Sushentsev N, et al. Excessive Media Consumption About COVID-19 is Associated With Increased State Anxiety: Outcomes of a Large Online Survey in Russia. J Med Internet Res 2020 Sep 11;22(9):e20955 [FREE Full text] [doi: 10.2196/20955] [Medline: 32788143]
15. Starcevic V, Berle D, Arnáez S. Recent Insights Into Cyberchondria. Curr Psychiatry Rep 2020 Aug 27;22(11):56 [FREE Full text] [doi: 10.1007/s11920-020-01179-8] [Medline: 32852626]
16. McManus F, Leung C, Muse K, Williams JMG. Understanding 'cyberchondria': an interpretive phenomenological analysis of the purpose, methods and impact of seeking health information online for those with health anxiety. Cogn Behav Therap 2015 Jan 12;7. [doi: 10.1017/S1754470X14000270]
17. van den Heuvel OA, Veale D, Stein DJ. Hypochondriasis: considerations for ICD-11. Braz J Psychiatry 2014;36 Suppl 1(suppl 1):21-27 [FREE Full text] [doi: 10.1590/1516-4446-2013-1218] [Medline: 25388609]
18. Maftei A, Holman AC. Cyberchondria During the Coronavirus Pandemic: The Effects of Neuroticism and Optimism. Front Psychol 2020 Oct 30;11:567345 [FREE Full text] [doi: 10.3389/fpsyg.2020.567345] [Medline: 33192848]
19. Gratz KL, Weiss NH, Tull MT. Examining Emotion Regulation as an Outcome, Mechanism, or Target of Psychological Treatments. Curr Opin Psychol 2015 Jun 01;3:85-90 [FREE Full text] [doi: 10.1016/j.copsyc.2015.02.010] [Medline: 25859561]
20. Nguyen LH, Drew DA, Graham MS, Joshi AD, Guo C, Ma W, COronavirus Pandemic Epidemiology Consortium. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. Lancet Public Health 2020 Sep;5(9):e475-e483 [FREE Full text] [doi: 10.1016/S2468-2667(20)30164-X] [Medline: 32745512]
21. Bendau A, Petzold MB, Pyrkosch L, Mascarell Maricic L, Betzler F, Rogoll J, et al. Associations between COVID-19 related media consumption and symptoms of anxiety, depression and COVID-19 related fear in the general population in Germany. Eur Arch Psychiatry Clin Neurosci 2021 Mar 20;271(2):283-291 [FREE Full text] [doi: 10.1007/s00406-020-01171-6] [Medline: 32691135]
22. Liu M, Zhang H, Huang H. Media exposure to COVID-19 information, risk perception, social and geographical proximity, and self-rated anxiety in China. BMC Public Health 2020 Nov 04;20(1):1649 [FREE Full text] [doi: 10.1186/s12889-020-09761-8] [Medline: 33148201]
23. Kubb C, Foran H. Measuring COVID-19 Related Anxiety in Parents: Psychometric Comparison of Four Different Inventories. JMIR Ment Health 2020 Dec 03;7(12):e24507 [FREE Full text] [doi: 10.2196/24507] [Medline: 33197233]
24. Mahmud MS, Talukder MU, Rahman SM. Does 'Fear of COVID-19' trigger future career anxiety? An empirical investigation considering depression from COVID-19 as a mediator. Int J Soc Psychiatry 2020 Jul 02 [FREE Full text] [doi: 10.1177/0020764020935488] [Medline: 32615833]
25. Arpaci I, Karataş K, Baloğlu M. The development and initial tests for the psychometric properties of the COVID-19 Phobia Scale (C19P-S). Pers Individ Dif 2020 Oct 01;164:110108 [FREE Full text] [doi: 10.1016/j.paid.2020.110108] [Medline: 32394993]
26. Feng L, Dong Z, Yan R, Wu X, Zhang L, Ma J, et al. Psychological distress in the shadow of the COVID-19 pandemic: Preliminary development of an assessment scale. Psychiatry Res 2020 Sep;291:113202 [FREE Full text] [doi: 10.1016/j.psychres.2020.113202] [Medline: 32535511]
27. Voitsidis P, Gliatas I, Bairachtari V, Papadopoulou K, Papageorgiou G, Parlapani E, et al. Insomnia during the COVID-19 pandemic in a Greek population. Psychiatry Res 2020 May 12;289:113076. [doi: 10.1016/j.psychres.2020.113076] [Medline: 33242799]
28. Menon V, Kar SK, Tripathi A, Nebhinani N, Varadharajan N. Cyberchondria: conceptual relation with health anxiety, assessment, management and prevention. Asian J Psychiatr 2020 Oct;53:102225. [doi: 10.1016/j.ajp.2020.102225] [Medline: 32585634]
29. McElroy E, Shevlin M. The development and initial validation of the cyberchondria severity scale (CSS). J Anxiety Disord 2014 Mar;28(2):259-265. [doi: 10.1016/j.janxdis.2013.12.007] [Medline: 24508033]
30. Bajcar B, Babiak J. Neuroticism and cyberchondria: The mediating role of intolerance of uncertainty and defensive pessimism. Personality and Individual Differences 2020 Aug;162:110006. [doi: 10.1016/j.paid.2020.110006]
31. Swee MB, Olino TM, Heimberg RG. Worry and anxiety account for unique variance in the relationship between intolerance of uncertainty and depression. Cogn Behav Ther 2019 May 25;48(3):253-264. [doi: 10.1080/16506073.2018.1533579] [Medline: 30354911]
32. Mertens G, Gerritsen L, Duijndam S, Salemink E, Engelhard IM. Fear of the coronavirus (COVID-19): Predictors in an online study conducted in March 2020. J Anxiety Disord 2020 Aug;74:102258 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102258] [Medline: 32569905]

33. Lee SA, Crunk EA. Fear and Psychopathology During the COVID-19 Crisis: Neuroticism, Hypochondriasis, Reassurance-Seeking, and Coronaphobia as Fear Factors. Omega (Westport) 2020 Aug 06:30222820949350. [doi: 10.1177/0030222820949350] [Medline: 32762291]
34. Taha S, Matheson K, Cronin T, Anisman H. Intolerance of uncertainty, appraisals, coping, and anxiety: the case of the 2009 H1N1 pandemic. Br J Health Psychol 2014 Sep 09;19(3):592-605. [doi: 10.1111/bjhp.12058] [Medline: 23834735]
35. Reiss S, Peterson RA, Gursky DM, McNally RJ. Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. Behaviour Research and Therapy 1986;24(1):1-8. [doi: 10.1016/0005-7967(86)90143-9]
36. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. The Lancet Psychiatry 2020 Jun 15;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
37. Fergus TA. Anxiety sensitivity and intolerance of uncertainty as potential risk factors for cyberchondria: A replication and extension examining dimensions of each construct. J Affect Disord 2015 Sep 15;184:305-309. [doi: 10.1016/j.jad.2015.06.017] [Medline: 26141807]
38. Fergus TA. Cyberchondria and intolerance of uncertainty: examining when individuals experience health anxiety in response to Internet searches for medical information. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):735-739. [doi: 10.1089/cyber.2012.0671] [Medline: 23992476]
39. Norr AM, Allan NP, Macatee RJ, Keough ME, Schmidt NB. The effects of an anxiety sensitivity intervention on anxiety, depression, and worry: mediation through affect tolerances. Behav Res Ther 2014 Aug;59:12-19 [FREE Full text] [doi: 10.1016/j.brat.2014.05.011] [Medline: 24949907]
40. McElroy E, Kearney M, Touhey J, Evans J, Cooke Y, Shevlin M. The CSS-12: Development and Validation of a Short-Form Version of the Cyberchondria Severity Scale. Cyberpsychol Behav Soc Netw 2019 May;22(5):330-335. [doi: 10.1089/cyber.2018.0624] [Medline: 31013440]
41. Ahorsu DK, Lin C, Imani V, Saffari M, Griffiths MD, Pakpour AH. The Fear of COVID-19 Scale: Development and Initial Validation. Int J Ment Health Addict 2020 Mar 27:1-9 [FREE Full text] [doi: 10.1007/s11469-020-00270-8] [Medline: 32226353]
42. Lee SA. Coronavirus Anxiety Scale: A brief mental health screener for COVID-19 related anxiety. Death Stud 2020;44(7):393-401. [doi: 10.1080/07481187.2020.1748481] [Medline: 32299304]
43. Taylor S, Zvolensky MJ, Cox BJ, Deacon B, Heimberg RG, Ledley DR, et al. Robust dimensions of anxiety sensitivity: development and initial validation of the Anxiety Sensitivity Index-3. Psychol Assess 2007 Jun;19(2):176-188. [doi: 10.1037/1040-3590.19.2.176] [Medline: 17563199]
44. Carleton RN, Norton MAPJ, Asmundson GJG. Fearing the unknown: a short version of the Intolerance of Uncertainty Scale. J Anxiety Disord 2007;21(1):105-117. [doi: 10.1016/j.janxdis.2006.03.014] [Medline: 16647833]
45. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/brm.41.4.1149]
46. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd Edition. New York, NY: Routledge; 1988.
47. Tabachnick B, Fidell L. Using Multivariate Statistics. Boston, MA: Pearson; 2014.
48. Hair J, Black W, Babin B, Anderson R, Black W, Anderson R. Chapter 4 Multivariate data analysis. In: Klemelä J, editor. Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons, Inc; 2018.
49. Kane L, Ashbaugh AR. Simple and parallel mediation: A tutorial exploring anxiety sensitivity, sensation seeking, and gender. TQMP 2017 Oct 01;13(3):148-165. [doi: 10.20982/tqmp.13.3.p148]
50. Hayes AF, Preacher KJ. Statistical mediation analysis with a multicategorical independent variable. Br J Math Stat Psychol 2014 Nov;67(3):451-470. [doi: 10.1111/bmsp.12028] [Medline: 24188158]
51. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol 2007;58:593-614 [FREE Full text] [doi: 10.1146/annurev.psych.58.110405.085542] [Medline: 16968208]
52. Lin C, Broström A, Griffiths MD, Pakpour AH. Investigating mediated effects of fear of COVID-19 and COVID-19 misunderstanding in the association between problematic social media use, psychological distress, and insomnia. Internet Interv 2020 Sep;21:100345 [FREE Full text] [doi: 10.1016/j.invent.2020.100345] [Medline: 32868992]
53. Liu Y, Cao L, Li X, Jia Y, Xia H. Awareness of mental health problems in patients with coronavirus disease 19 (COVID-19): A lesson from an adult man attempting suicide. Asian J Psychiatr 2020 Jun;51:102106 [FREE Full text] [doi: 10.1016/j.ajp.2020.102106] [Medline: 32334411]
54. Jungmann SM, Witthöft M. Health anxiety, cyberchondria, and coping in the current COVID-19 pandemic: Which factors are related to coronavirus anxiety? J Anxiety Disord 2020 Jun;73:102239 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102239] [Medline: 32502806]
55. Wheaton MG, Messner GR, Marks JB. Intolerance of uncertainty as a factor linking obsessive-compulsive symptoms, health anxiety and concerns about the spread of the novel coronavirus (COVID-19) in the United States. J Obsessive Compuls Relat Disord 2021 Jan;28:100605 [FREE Full text] [doi: 10.1016/j.jocrd.2020.100605] [Medline: 33251098]
56. Tull MT, Barbano AC, Scamaldo KM, Richmond JR, Edmonds KA, Rose JP, et al. The prospective influence of COVID-19 affective risk assessments and intolerance of uncertainty on later dimensions of health anxiety. J Anxiety Disord 2020 Oct;75:102290 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102290] [Medline: 32823216]

57. Seyed Hashemi SG, Hosseinnezhad S, Dini S, Griffiths MD, Lin C, Pakpour AH. The mediating effect of the cyberchondria and anxiety sensitivity in the association between problematic internet use, metacognition beliefs, and fear of COVID-19 among Iranian online population. Heliyon 2020 Oct;6(10):e05135 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e05135] [Medline: 33072911]
58. Freeston M, Tiplady A, Mawn L, Bottesi G, Thwaites S. Towards a model of uncertainty distress in the context of Coronavirus (COVID-19). PsyArXiv Preprints. Preprint posted online April 14, 2020 2021. [doi: 10.31234/osf.io/v8q6m]
59. Bottesi G, Noventa S, Freeston MH, Ghisi M. Seeking certainty about Intolerance of Uncertainty: Addressing old and new issues through the Intolerance of Uncertainty Scale-Revised. PLoS One 2019;14(2):e0211929 [FREE Full text] [doi: 10.1371/journal.pone.0211929] [Medline: 30742694]
60. Gutierrez R, Hirani T, Curtis L, Ludlow AK. Metacognitive beliefs mediate the relationship between anxiety sensitivity and traits of obsessive-compulsive symptoms. BMC Psychol 2020 Apr 26;8(1):40 [FREE Full text] [doi: 10.1186/s40359-020-00412-6] [Medline: 32336292]
61. Oser M, Khan A, Kolodziej M, Gruner G, Barsky AJ, Epstein L. Mindfulness and Interoceptive Exposure Therapy for Anxiety Sensitivity in Atrial Fibrillation: A Pilot Study. Behav Modif 2021 May 24;45(3):462-479. [doi: 10.1177/0145445519877619] [Medline: 31550903]
62. Naragon-Gainey K. Meta-analysis of the relations of anxiety sensitivity to the depressive and anxiety disorders. Psychol Bull 2010 Jan;136(1):128-150. [doi: 10.1037/a0018055] [Medline: 20063929]
63. Hovenkamp-Hermelink J, Van Der Veen D, Oude Voshaar R, Batelaan N, Penninx B, Schoevers R, et al. Anxiety sensitivity: longitudinal stability and association with anxiety severity. Front. Psychiatry 2017;8. [doi: 10.3389/conf.fpsyt.2017.48.00039]
64. Stanley IH, Boffa JW, Rogers ML, Hom MA, Albanese BJ, Chu C, et al. Anxiety sensitivity and suicidal ideation/suicide risk: A meta-analysis. J Consult Clin Psychol 2018 Nov;86(11):946-960 [FREE Full text] [doi: 10.1037/ccp0000342] [Medline: 30335426]
65. Barke A, Bleichhardt G, Rief W, Doering BK. The Cyberchondria Severity Scale (CSS): German Validation and Development of a Short Form. Int J Behav Med 2016 Oct 1;23(5):595-605. [doi: 10.1007/s12529-016-9549-8] [Medline: 26931780]
66. Nazari N, Griffiths MD. Psychometric validation of the Persian version of the Emotional Style Questionnaire. Curr Psychol 2020 Nov 26:1-13 [FREE Full text] [doi: 10.1007/s12144-020-01205-1] [Medline: 33262561]
67. Shi L, Lu Z, Que J, Huang X, Liu L, Ran M, et al. Prevalence of and Risk Factors Associated With Mental Health Symptoms Among the General Population in China During the Coronavirus Disease 2019 Pandemic. JAMA Netw Open 2020 Jul 01;3(7):e2014053 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.14053] [Medline: 32609353]
68. Lebel C, MacKinnon A, Bagshawe M, Tomfohr-Madsen L, Giesbrecht G. Elevated depression and anxiety symptoms among pregnant individuals during the COVID-19 pandemic. J Affect Disord 2020 Dec 01;277:5-13 [FREE Full text] [doi: 10.1016/j.jad.2020.07.126] [Medline: 32777604]
69. Blüml V, Kapusta ND, Doering S, Brähler E, Wagner B, Kersting A. Personality factors and suicide risk in a representative sample of the German general population. PLoS One 2013 Oct 4;8(10):e76646 [FREE Full text] [doi: 10.1371/journal.pone.0076646] [Medline: 24124582]
70. Durak Batıgün A, Şenkal Ertürk İ, Gör N, Kömürcü Akik B. The pathways from distress tolerance to Cyberchondria: A multiple-group path model of young and middle adulthood samples. Curr Psychol 2020 Sep 05:1-9 [FREE Full text] [doi: 10.1007/s12144-020-01038-y] [Medline: 32921966]
71. Fergus TA, Spada MM. Cyberchondria: Examining relations with problematic Internet use and metacognitive beliefs. Clin Psychol Psychother 2017 Nov;24(6):1322-1330. [doi: 10.1002/cpp.2102] [Medline: 28621035]
72. Griffiths MD. The Use of Online Methodologies in Data Collection for Gambling and Gaming Addictions. Int J Ment Health Addiction 2009 Apr 8;8(1):8-20. [doi: 10.1007/s11469-009-9209-1]
73. Sauer-Zavala S, Bentley KH, Steele SJ, Tirpak JW, Ametaj AA, Nauphal M, et al. Treating depressive disorders with the Unified Protocol: A preliminary randomized evaluation. J Affect Disord 2020 Mar 01;264:438-445 [FREE Full text] [doi: 10.1016/j.jad.2019.11.072] [Medline: 31759672]
74. Nazari N, Sadeghi M, Ghadampour E, Mirzaeefar D. Transdiagnostic treatment of emotional disorders in people with multiple sclerosis: randomized controlled trial. BMC Psychol 2020 Oct 31;8(1):114 [FREE Full text] [doi: 10.1186/s40359-020-00480-8] [Medline: 33129356]
75. Restubog SLD, Ocampo ACG, Wang L. Taking control amidst the chaos: Emotion regulation during the COVID-19 pandemic. J Vocat Behav 2020 Jun;119:103440 [FREE Full text] [doi: 10.1016/j.jvb.2020.103440] [Medline: 32390659]
76. Holmes EA, Ghaderi A, Harmer CJ, Ramchandani PG, Cuijpers P, Morrison AP, et al. The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. The Lancet Psychiatry 2018 Mar;5(3):237-286. [doi: 10.1016/S2215-0366(17)30513-8] [Medline: 29482764]
77. Jokic-Begic N, Lauri Korajlija A, Mikac U. Cyberchondria in the age of COVID-19. PLoS One 2020;15(12):e0243704 [FREE Full text] [doi: 10.1371/journal.pone.0243704] [Medline: 33332400]
78. Bonanno GA, Maccallum F, Malgaroli M, Hou WK. The Context Sensitivity Index (CSI): Measuring the Ability to Identify the Presence and Absence of Stressor Context Cues. Assessment 2020 Mar 21;27(2):261-273. [doi: 10.1177/1073191118820131] [Medline: 30577707]

79. Harvey MM, Coifman KG, Ross G, Kleinert D, Giardina P. Contextually appropriate emotional word use predicts adaptive health behavior: Emotion context sensitivity and treatment adherence. J Health Psychol 2016 May 06;21(5):579-589. [doi: 10.1177/1359105314532152] [Medline: 24801328]
80. Nazari N, Aligholipour A, Sadeghi M. Transdiagnostic treatment of emotional disorders for women with multiple sclerosis: a randomized controlled trial. BMC Womens Health 2020 Oct 31;20(1):245 [FREE Full text] [doi: 10.1186/s12905-020-01109-z] [Medline: 33129298]

Abbreviations

ASI-3: Anxiety Sensitivity Index-3
CAS: Coronavirus Anxiety Scale
CSS-12: Cyberchondria Severity Scale–Short Form
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, fifth edition
FCV-19S: Fear of COVID-19 Scale
IUS-12: Intolerance of Uncertainty Scale-12
SARS: severe acute respiratory syndrome
SEM: structural equation modeling
VIF: variance inflation factor

Edited by G Eysenbach; submitted 26.12.20; peer-reviewed by L Feng, A Chatterjee, D Manavipoor; comments to author 13.02.21; revised version received 18.02.21; accepted 03.05.21; published 09.06.21.

Please cite as:
Wu X, Nazari N, Griffiths MD
Using Fear and Anxiety Related to COVID-19 to Predict Cyberchondria: Cross-sectional Survey Study
J Med Internet Res 2021;23(6):e26285
URL: https://www.jmir.org/2021/6/e26285
doi:10.2196/26285
PMID:34014833

Original Paper

Assessing COVID-19 Vaccine Hesitancy, Confidence, and Public Engagement: A Global Social Listening Study

Zhiyuan Hou[1,2,3*], PhD; Yixin Tong[1*], BSc; Fanxing Du[1], MPH; Linyao Lu[1], BSc; Sihong Zhao[1], BSc; Kexin Yu[1], BSc; Simon J Piatek[4], PhD; Heidi J Larson[4], PhD; Leesa Lin[4,5], PhD

[1]School of Public Health, Fudan University, Shanghai, China
[2]National Health Commission Key Laboratory of Health Technology Assessment, Fudan University, Shanghai, China
[3]Global Health Institute, Fudan University, Shanghai, China
[4]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom
[5]Laboratory of Data Discovery for Health, Hong Kong Science Park, Hong Kong SAR, China
*these authors contributed equally

Corresponding Author:
Zhiyuan Hou, PhD
School of Public Health
Fudan University
Mailbox 250, 138# Yixueyuan Road, Xuhui District
Shanghai, 200030
China
Phone: 86 2133563935
Email: zyhou@fudan.edu.cn

Abstract

Background: Monitoring public confidence and hesitancy is crucial for the COVID-19 vaccine rollout. Social media listening (infoveillance) can not only monitor public attitudes on COVID-19 vaccines but also assess the dissemination of and public engagement with these opinions.

Objective: This study aims to assess global hesitancy, confidence, and public engagement toward COVID-19 vaccination.

Methods: We collected posts mentioning the COVID-19 vaccine between June and July 2020 on Twitter from New York (United States), London (United Kingdom), Mumbai (India), and Sao Paulo (Brazil), and Sina Weibo posts from Beijing (China). In total, we manually coded 12,886 posts from the five global metropolises with high COVID-19 burdens, and after assessment, 7032 posts were included in the analysis. We manually double-coded these posts using a coding framework developed according to the World Health Organization's Confidence, Complacency, and Convenience model of vaccine hesitancy, and conducted engagement analysis to investigate public communication about COVID-19 vaccines on social media.

Results: Among social media users, 36.4% (571/1568) in New York, 51.3% (738/1440) in London, 67.3% (144/214) in Sao Paulo, 69.8% (726/1040) in Mumbai, and 76.8% (2128/2770) in Beijing indicated that they intended to accept a COVID-19 vaccination. With a high perceived risk of getting COVID-19, more tweeters in New York and London expressed a lack of confidence in vaccine safety, distrust in governments and experts, and widespread misinformation or rumors. Tweeters from Mumbai, Sao Paulo, and Beijing worried more about vaccine production and supply, whereas tweeters from New York and London had more concerns about vaccine distribution and inequity. Negative tweets expressing lack of vaccine confidence and misinformation or rumors had more followers and attracted more public engagement online.

Conclusions: COVID-19 vaccine hesitancy is prevalent worldwide, and negative tweets attract higher engagement on social media. It is urgent to develop an effective vaccine campaign that boosts public confidence and addresses hesitancy for COVID-19 vaccine rollouts.

(J Med Internet Res 2021;23(6):e27632) doi:10.2196/27632

KEYWORDS

COVID-19 vaccine; hesitancy; infoveillance; infodemiology; confidence; acceptance; engagement; social media; COVID-19

Introduction

As of January 2021, the COVID-19 pandemic has led to more than 100 million cases and 2 million deaths worldwide [1]. Although personal prevention measures such as mask wearing and social distancing have been shown to be effective in curbing the spread of COVID-19 [2], vaccination is expected to be the key to the long-term prevention and control of the pandemic [3,4]. The COVID-19 pandemic has triggered intense global research and development (R&D) of vaccines against the disease. Several candidate vaccines advanced to Phase III clinical trials in mid-2020, including the Oxford/AstraZeneca, Sinopharm, Sinovac, BioNTech/Pfizer, and Moderna vaccines, and had finished clinical trials at the end of 2020. These vaccines have been approved for use by December 2020 in some countries such as the United Kingdom, the United States, and China [5]. To ensure universal vaccination coverage, governments must enhance public confidence, address the issue of vaccine hesitancy, and design community engagement strategies for COVID-19 vaccine rollouts.

Although immunization has proved successful in reducing the global burden of illness and death, a range of concerns have converged to affect public confidence in vaccines. When vaccine confidence breaks down, hesitancy can cause serious consequences such as delays, refusals, and disruptions to research and delivery programs, and sometimes leads to the resurgence of disease outbreaks [6]. Vaccine hesitancy has proliferated over the decades and was cited by the World Health Organization (WHO) as one of the top 10 global health threats in 2019 [7,8]. Vaccine hesitancy is complex, multifactorial, and influenced by a combination of emotional, cultural, social, spiritual, and political factors. It can vary across countries, vaccines, and time. Reports have indicated that hesitancy toward general vaccines is prevalent among caregivers of children worldwide, at rates such as 45.8% in France (2016) [7], 31.8% in the United States (2014) [9], 24.6% in Italy (2017) [10], and 23% in Brazil (2016) [11]. The prevalence of vaccine hesitancy among health care providers ranges from 2% to 16% across different countries [12]. The accelerated R&D process of the COVID-19 vaccine may further exacerbate public concern on its safety and effectiveness [13]. Similarly, the novelty of the disease, politicization of the vaccine, and distrust in experts and governments have increased uncertainty about COVID-19 vaccination [3]. Therefore, it is necessary to assess public confidence and acceptance toward the COVID-19 vaccine to prepare for vaccine introduction.

Social media has become a source of data for detecting outbreaks and understanding public attitudes and behaviors during public health emergencies [14-16]. Large amounts of real-time data posted on social media platforms can be used to quickly identify public attitudes on COVID-19 vaccines as a way to support health communication and health promotion messaging. A growing body of literature has used social media platforms such as Twitter and Facebook for public health research [15]. Compared with traditional surveys, social media listening can not only monitor public attitudes in a timely manner but also assess the dissemination of and public engagement with these opinions [17]. Individuals are increasingly using social media to communicate with each other, and public engagement can assess how various messages around COVID-19 vaccines spread on social media. Although several previous studies investigated public acceptance toward COVID-19 vaccines using questionnaire surveys, it is unknown how these opinions spread among the public.

By using social media listening data from the largest metropolises worldwide, this study aimed to assess global vaccine hesitancy and confidence toward COVID-19 vaccination and public engagement in communications about COVID-19 vaccines.

Methods

Data Collection

Twitter is one of the most popular social media platforms in the world; Sina Weibo, Twitter in China, is the most influential social media platform in China, with over 500 million users. Users can share information or opinions by tweets or posts on these platforms. Using the Meltwater platform [18], we collected posts mentioning COVID-19 vaccines on Twitter from New York (United States), London (United Kingdom), Mumbai (India), Sao Paulo (Brazil), and Sina Weibo posts from Beijing (China). The five metropolises were selected due to their high disease burden of COVID-19. The data covered the period from June 13 to July 31, 2020, when five COVID-19 vaccines started their Phase III clinical trials worldwide, including the Oxford/AstraZeneca, Sinopharm, Sinovac, BioNTech/Pfizer, and Moderna vaccines [19]. The keywords included “COVID vaccin*” OR “COVID-19 vaccin*” OR “COVID19 vaccin*” OR “coronavirus vaccin*” OR “vaccin* for coronavirus” OR “vaccin* for COVID.” Each post record comprised account name, contents, post time, the number of followers, and engagement data. Since our study aimed to assess public attitudes toward COVID-19 vaccination, we only included tweets or Weibo posts from individual accounts and excluded those from news and organizational accounts. Duplicate tweets, tweets with identical text but different tweet identifications, retweets, quotes without comments, and irrelevant tweets were removed [20]. This study was exempt from ethical review because it examined retrospective publicly available data.

Content Analysis

We identified and classified posts describing personal opinions or discussion of COVID-19 vaccines from Twitter and Sina Weibo. A coding framework (Multimedia Appendix 1) was developed for content analysis according to the WHO’s Confidence, Complacency, and Convenience (“3 Cs”) model of vaccine hesitancy [21] and validated through manual annotation of the subset with 500 posts. All posts were double-coded independently, and a third coder resolved disagreements. Posts were initially classified as relevant or irrelevant to personal opinions toward COVID-19 vaccines, and relevant posts were further classified to the predefined categories in Multimedia Appendix 1. Predefined categories included attitudes toward COVID-19 vaccination (accept, neutral, doubt, or refuse), expectations of COVID-19 vaccine R&D and introduction (positive, neutral, or negative), confidence in COVID-19 vaccine importance (important or not), confidence in COVID-19 vaccine effectiveness (effective or not), confidence in COVID-19 vaccine safety (safe or not), trust in governments (trust or not), trust in experts (trust or not), misinformation or rumors about all vaccines, complacency (perceived risk of getting COVID-19: high or low), COVID-19 vaccine convenience (accessibility, distribution, or affordability), COVID-19 vaccine types (AstraZeneca, Moderna, Pfizer, or Chinese vaccines), and others. Each post could be classified as one category, multiple categories, or no category. We described simple counts and percentages of posts for each topic on COVID-19 vaccination.

Social Media Engagement Analysis

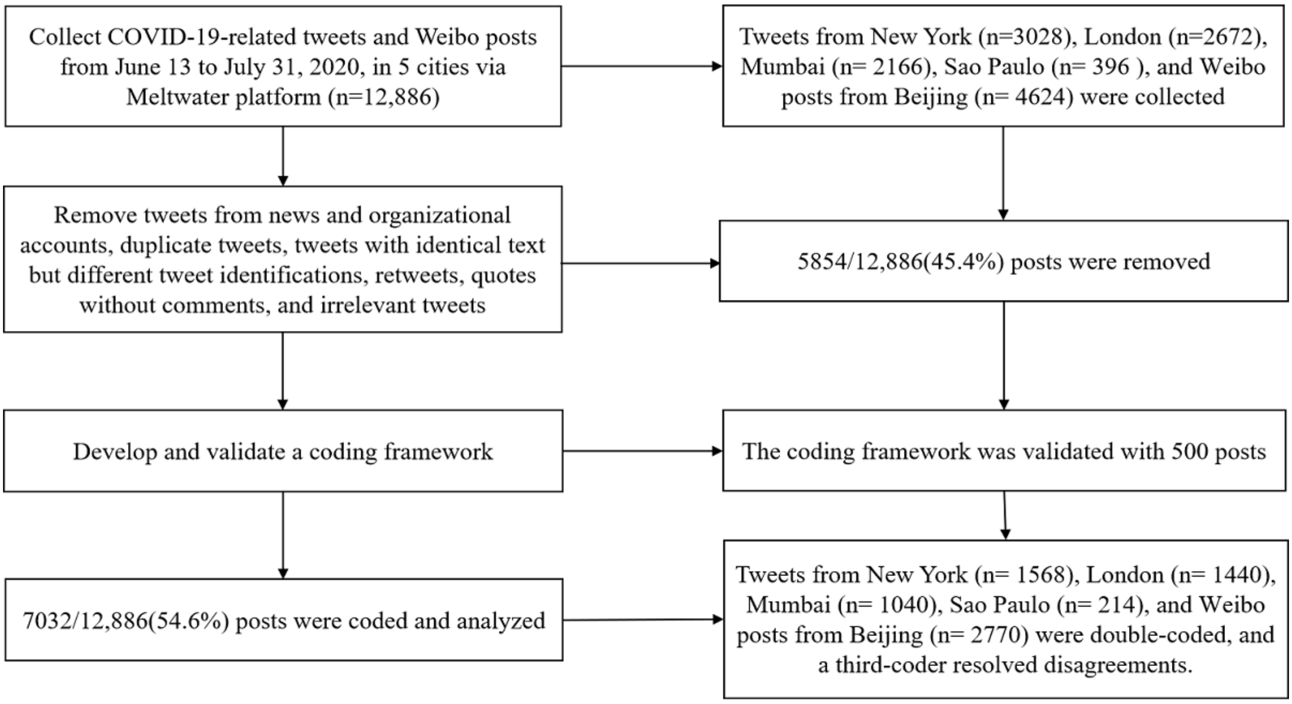
We also conducted social media engagement analysis to investigate public communication and interaction about various topics relating to the COVID-19 vaccine online. In Twitter, the engagement metric measures all actions viewers have taken as a result of seeing tweets and engaging with tweets [22]. At the level of public engagement and interaction, Twitter engagement refers to retweets, follows, replies, favorites, and clicks on the tweets. It covers the three metrics of social media engagement: popularity based on the number of likes for tweets, commitment based on the number of comments for tweets, and virality based on the number of shares for tweets [22]. The mean number and SD of engagement with and followers of tweeters were presented for each topic of COVID-19 vaccines.

Results

Description of Analyzed Social Media Posts

Figure 1 shows the process of data selection and analysis. In total, we collected 12,886 social media posts mentioning a COVID-19 vaccine. There were 3028 tweets on Twitter from New York, 2672 tweets from London, 2166 tweets from Mumbai, 396 tweets from Sao Paulo, and 4624 Sina Weibo posts from Beijing. After assessment, 7032 posts met the inclusion criteria and were included in our content analysis, including 1568 tweets from New York, 1440 tweets from London, 1040 tweets from Mumbai, 214 tweets from Sao Paulo, and 2770 Sina Weibo posts from Beijing.

Figure 1. Flowchart of data process and analysis.



COVID-19 Vaccine Hesitancy and Confidence

Figure 2 presents social media users' attitudes toward COVID-19 vaccination. Among social media users, 36.4% (571/1568) in New York and 51.3% (738/1440) in London reported willingness to accept COVID-19 vaccines, much lower than observed in metropolises in lower- and middle-income countries (LMICs), such as 69.8% (726/1040) in Mumbai, 67.3% (144/214) in Sao Paulo, and 76.8% (2128/2770) in Beijing. Although about 10% to 20% of users doubted the safety or effectiveness of COVID-19 vaccines in each metropolis, 20% (313/1568) in New York and 15.1% (218/1440) in London expressed refusal, much higher than other metropolises (<5%), leading to prevalent vaccine hesitancy. About 50% to 80% of users discussed the R&D and introduction of the COVID-19 vaccine globally, with most showing positive expectations, and the remaining users did not mention the R&D and introduction of COVID-19 vaccines (Figure 3). The discussion levels on the R&D and introduction of the COVID-19 vaccine in New York and London were lower than those in Mumbai, Beijing, and Sao Paulo.

Figure 2. Social media users' attitudes toward COVID-19 vaccination.

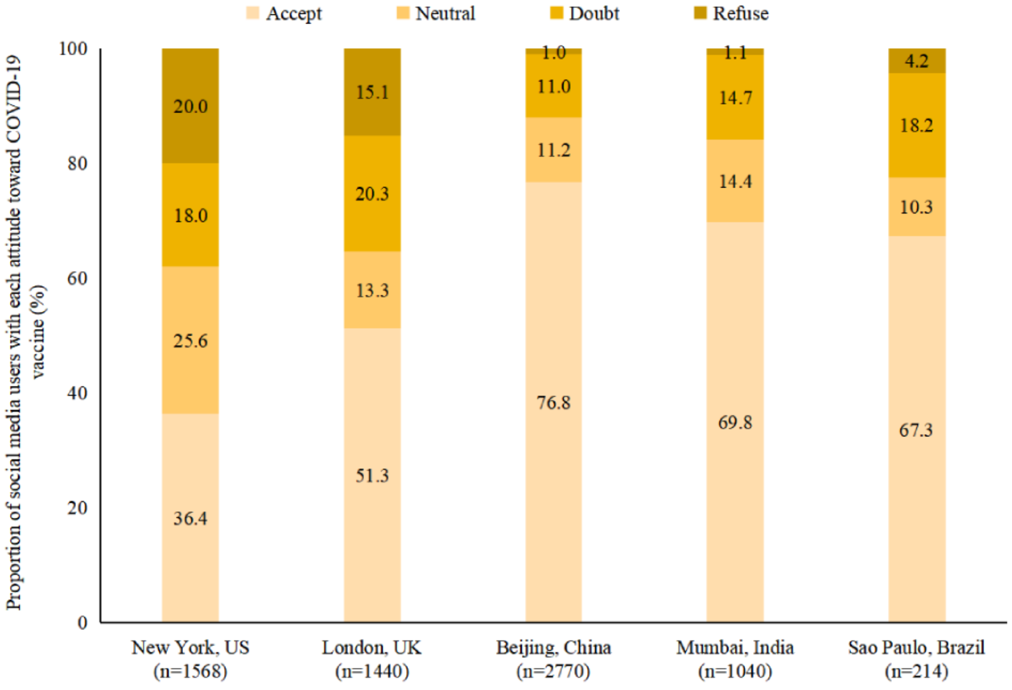


Figure 3. Expectations of COVID-19 vaccine R&D and introduction. R&D: research and development.

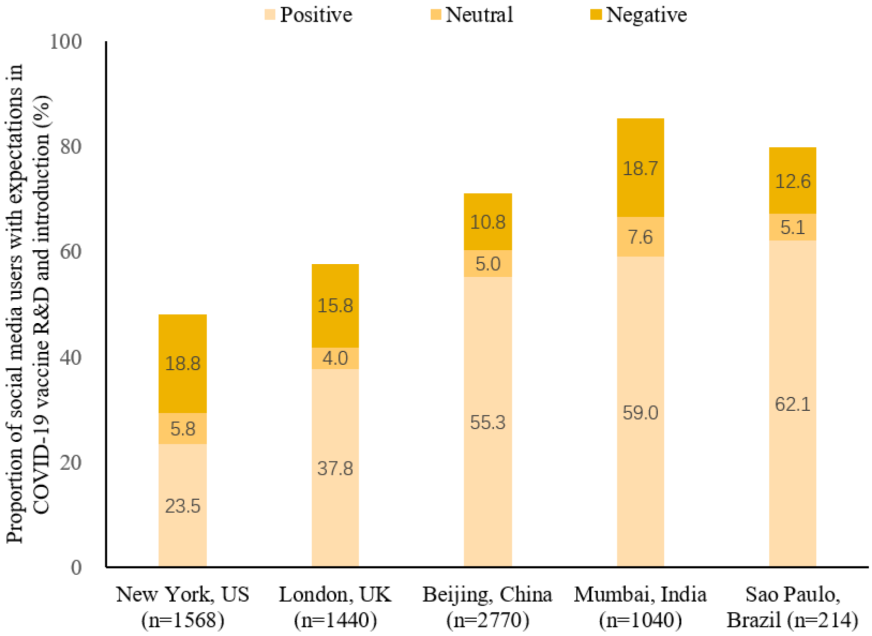


Table 1 conveys confidence, complacency, and convenience related to COVID-19 vaccination on social media. Although most tweets in each metropolis regarded COVID-19 vaccination as important and effective, proportions of tweets perceiving COVID-19 vaccines as unsafe in high-income countries (HICs; 133/1568, 8.5% in New York and 157/1440, 10.9% in London) were much higher than those in other metropolises (<4%). These HIC users expressed distrust in governments (>10%) and experts (around 5%). Furthermore, 11.8% (185/1568) and 11% (158/1440) of users in New York and London, respectively, mentioned misinformation, rumors, or antivaccine campaigns, but few users in LMIC metropolises (<7%) shared these concerns.

Table 1. Confidence, complacency, and convenience related to COVID-19 vaccines on social media in 2020.

Topics	New York, US	London, UK	Beijing, China	Mumbai, India	Sao Paulo, Brazil
Total posts[a], n	3028	2672	4624	2166	396
Irrelevant posts	1460	1232	1854	1126	182
Relevant posts	1568	1440	2770	1040	214
Topics among relevant posts[b], n (%)					
Vaccine confidence					
Perceived importance of vaccines	115 (7.3)	130 (9.0)	712 (25.7)	140 (13.5)	22 (10.3)
Important	96 (6.1)	115 (8.0)	651 (23.5)	132 (12.7)	22 (10.3)
Unimportant	19 (1.2)	16 (1.1)	61 (2.2)	8 (0.8)	0 (0.0)
Perceived effectiveness of vaccines	217 (13.8)	361 (25.1)	928 (33.5)	294 (28.3)	71 (33.2)
Effective	142 (9.1)	305 (21.2)	674 (24.3)	261 (25.1)	68 (31.8)
Ineffective	75 (4.8)	56 (3.9)	254 (9.2)	32 (3.1)	3 (1.4)
Perceived safety of vaccines	168 (10.7)	354 (24.6)	201 (7.3)	190 (18.3)	47 (22.0)
Safe	35 (2.2)	197 (13.7)	143 (5.2)	149 (14.3)	43 (20.1)
Unsafe	133 (8.5)	157 (10.9)	58 (2.1)	41 (3.9)	4 (1.9)
Trust in governments	294 (18.8)	177 (12.3)	148 (5.3)	72 (6.9)	8 (3.7)
Trust	14 (0.9)	19 (1.3)	138 (5.0)	16 (1.5)	1 (0.5)
Distrust	280 (17.9)	158 (11.0)	10 (0.4)	56 (5.4)	7 (3.3)
Trust in experts	90 (5.7)	111 (7.7)	—[c]	78 (7.5)	2 (0.9)
Trust	14 (0.9)	44 (3.1)	—	24 (2.3)	2 (0.9)
Distrust	76 (4.8)	67 (4.7)	—	54 (5.2)	0 (0.0)
Information around vaccines					
Misinformation or rumors	185 (11.8)	158 (11.0)	188 (6.8)	30 (2.9)	12 (5.6)
Complacency					
Perceived risk of getting COVID-19	89 (5.7)	52 (3.6)	742 (26.8)	45 (4.3)	4 (1.9)
High	80 (5.1)	45 (3.1)	688 (24.8)	41 (3.9)	4 (1.9)
Low	9 (0.6)	7 (0.5)	54 (1.9)	4 (0.4)	0 (0.0)
Vaccine convenience					
Vaccine accessibility	94 (6.0)	76 (5.3)	283 (10.2)	99 (9.5)	49 (22.9)
Vaccine distribution	309 (19.7)	261 (18.1)	325 (11.7)	63 (6.1)	26 (12.1)
Vaccine affordability	107 (6.8)	40 (2.8)	197 (7.1)	33 (3.2)	20 (9.3)
Vaccine types	259 (16.5)	470 (32.6)	941 (34.0)	270 (26.0)	122 (57.0)
AstraZeneca	102 (6.5)	388 (26.9)	82 (3.0)	218 (21.0)	63 (29.4)
Moderna	51 (3.3)	30 (2.1)	37 (1.3)	31 (3.0)	21 (9.8)
Pfizer	88 (5.6)	26 (1.8)	34 (1.2)	18 (1.7)	27 (12.6)
Chinese vaccines	25 (1.6)	26 (1.8)	831 (30.0)	19 (1.8)	13 (6.1)
Others	83 (5.3)	51 (3.5)	990 (35.7)	34 (3.3)	5 (2.3)

[a] We assessed 50% random samples from New York and London due to the large sample size, and full samples in Beijing, Mumbai, and Sao Paulo.

[b] Topics are calculated among relevant posts.

[c] Trust in experts is not measured specifically for Sina Weibo posts from Beijing.

Regarding complacency, most social media users perceived the risk of getting COVID-19 to be high despite only a small number of them (around 5% except Beijing) directly mentioning the risk of getting COVID-19. For vaccination convenience, approximately 5% of HIC users mentioned vaccine accessibility, whereas more users in LMIC metropolises (10%-20%) worried about vaccine accessibility, including production and supply capacity for COVID-19 vaccines. Nearly 20% of HIC users worried about vaccine distribution, including priority vaccination groups, whereas less users in LMIC metropolises (around 10%) mentioned this subject. Few users (about 5%) considered vaccine affordability in mid-2020. There were differential concerns on vaccine accessibility worldwide.

The proportion of tweets mentioning specific vaccine types was 16.5% (259/1568) in New York, 57% (122/214) in Sao Paulo, and around 30% in the remaining three metropolises. Tweets in New York mainly discussed COVID-19 vaccines produced in the United States, and tweets in other metropolises discussed vaccines produced worldwide. In terms of specific COVID-19 vaccines that had started Phase III clinical trials, social media users in metropolises except for Beijing mostly mentioned the Oxford/AstraZeneca vaccine, followed by the BioNTech/Pfizer and Moderna vaccines; the Chinese vaccine was the least frequently mentioned. This indicated that public attention was consistent to the development progress of specific vaccines.

Engagement Levels of COVID-19 Vaccine–Related Tweets

Table 2 presents the engagement metrics of COVID-19 vaccine–related tweets by topic. On average, tweeters who posted COVID-19 vaccine–related tweets had 3634 followers, and their followers ranged from 1637 (expressing COVID-19 vaccination is important) to 4778 followers (expressing COVID-19 vaccination is unsafe) and 7430 followers (expressing COVID-19 vaccination is ineffective). Overall, tweeters expressing negative attitudes toward COVID-19 vaccines, such as doubt or refusal to vaccinate, lack of confidence in vaccines (importance, effectiveness, and safety), and misinformation or rumors, had more followers than those expressing positive attitudes.

COVID-19 vaccine–related tweets attracted 6.1 engagement values on average. When comparing engagement levels on different topics regarding COVID-19 vaccines, we found that the most active topics were misinformation (engagement value 14.0), followed by confidence in vaccine safety (13.1), vaccine effectiveness (10.9), and governments (9.7), whereas the least active topics were vaccine convenience (2.6-5.8), confidence in vaccine importance (3.6) and experts (3.8), and complacency (5.4). For topics on vaccine hesitancy and expectations in vaccine R&D and introduction, tweets with clear attitudes attracted more engagement than neutral tweets, but there was little difference in public engagement between tweets with positive and negative attitudes. For each topic regarding vaccine confidence except distrust in governments, lack of confidence propagated more than having confidence, especially for vaccine safety (16.7 engagement for tweets expressing vaccine is unsafe vs 10.3 engagement for tweets expressing vaccine is safe); trust in governments attracted much more engagement than distrust (52.8 vs 5.4), representing the public's common expectations of governments taking an active role in controlling the epidemic.

Table 2. Followers and engagements of COVID-19 vaccine–related tweets in 2020.

Topics	Followers, mean (SD)	Engagements, mean (SD)
Vaccine hesitancy		
Attitudes toward COVID-19 vaccination	3633.6 (12,614.4)	6.1 (61.4)
Accept	2448.8 (8614.6)	7.9 (93.3)
Neutral	4333.0 (16,250.7)	5.8 (28.5)
Doubt	4593.3 (18,245.7)	10.0 (93.4)
Refuse	2708.4 (12,504.3)	6.4 (35.4)
Expectations of COVID-19 vaccine R&D[a] and introduction	3011.0 (11,759.4)	8.4 (94.4)
Positive	2527.7 (9083.2)	8.8 (104.3)
Neutral	4449.8 (16,399.9)	5.9 (31.9)
Negative	3629.4 (14,833.2)	8.2 (84.0)
Vaccine confidence		
Perceived importance of vaccines	1763.6 (3727.9)	3.6 (10.7)
Important	1636.5 (3167.2)	3.4 (9.9)
Unimportant	2843.1 (6811.7)	5.4 (16.5)
Perceived effectiveness of vaccines	3373.4 (13,960.7)	10.9 (115.5)
Effective	2505.6 (8366.8)	10.6 (119.5)
Ineffective	7430.1 (27,615.6)	12.2 (94.7)
Perceived safety of vaccines	3274.3 (14,058.0)	13.1 (137.2)
Safe	2086.2 (6191.7)	10.3 (137.4)
Unsafe	4778.0 (19,897.0)	16.7 (137.1)
Trust in governments	3865.2 (16,363.8)	9.7 (111.8)
Trust	3055.6 (6573.5)	52.8 (360.4)
Distrust	3946.0 (17,036.6)	5.4 (28.7)
Trust in experts	1887.9 (4812.1)	3.8 (14.7)
Trust	2029.1 (4076.9)	3.0 (6.5)
Distrust	1827.7 (5102.0)	4.1 (17.1)
Information around vaccines		
Misinformation or rumors	3127.7 (11,610.5)	14.0 (121.2)
Vaccine complacency		
Perceived risk of getting COVID-19	3017.9 (8282.9)	5.4 (15.6)
High	2909.7 (7848.9)	5.1 (15.5)
Low	3937.6 (11,554.1)	8.1 (16.5)
Vaccine convenience		
Vaccine accessibility	2351.5 (8915.4)	2.6 (7.8)
Vaccine distribution	3531.5 (12,541.6)	5.8 (26.2)
Vaccine affordability	2525.3 (4895.6)	3.1 (8.3)

[a] R&D: research and development.

Discussion

Principal Findings

This social listening study in large metropolitan areas in five countries examined hesitancy toward COVID-19 vaccination; perceptions of vaccine confidence, complacency, and convenience; and level of online public engagement. We found that COVID-19 vaccine hesitancy was prevalent worldwide, with the highest prevalence in New York and London, followed by Sao Paulo and Mumbai, and the lowest in Beijing. With high perceived risk of getting COVID-19, social media users in HICs, including the United States and the United Kingdom, expressed low acceptance of COVID-19 vaccines, serious concerns regarding vaccine safety, and distrust in governments and experts. There were different concerns about vaccination convenience and accessibility between HICs and LMICs. Negative tweets expressing lack of vaccine confidence and misinformation or rumors had more followers and attracted more online public engagement.

Overall, social media users expressed relatively high hesitancy toward COVID-19 vaccination across metropolises with high COVID-19 burdens. There were no metropolises where willingness to accept a COVID-19 vaccine exceeded 80%, and only between 36% to 50% of tweeters accepted the COVID-19 vaccine especially in New York and London. These social listening results were similar to global survey studies [13,23], which showed that COVID-19 vaccine acceptance was higher in China, India, and Brazil but lower in the United States and the United Kingdom. Over the past decade, vaccine refusal has accelerated worldwide, and antivaccine activities have been amplified through political activities and social media [24]. The prevalent hesitancy toward COVID-19 vaccines could potentially lead to low vaccination coverage, which will further delay global control of the pandemic and societal and economic recovery. Recent studies estimated a COVID-19 R_0 of around 3.87 for Europe and around 3.45 for the United States [25,26], implying herd immunity thresholds of 74% and 71%, respectively. The current acceptance levels among tweeters in New York, London, Mumbai, and Sao Paulo is insufficient to reach this 71% to 74% threshold; one exception is Beijing with an acceptance level at 76.8%. As governments are preparing to introduce COVID-19 vaccines and initiate postpandemic recovery, the need to develop an effective vaccine campaign for the rollout that boosts public confidence and addresses hesitancy is urgent.

A previous study showed that confidence in vaccines and governments was strongly associated with vaccine acceptance and uptake [13]. However, our findings showed lack of confidence in vaccine safety, distrust in governments and experts, and widespread misinformation or rumors, especially in HICs. Public confidence in the safety and effectiveness of COVID-19 vaccines was far lower than the confidence level in general vaccines. In the 2018 Global Monitor Survey, 79% and 84% of the public agreed that vaccines were safe and effective globally [27]. During COVID-19, the development of a vaccine for a new pathogen has been pushed much faster than ever before, and new bioscience technologies (eg, mRNA vaccine) are being used in humans for the first time [5,28]. In light of the accelerated R&D process, any negative news related to vaccine failure may weaken public confidence in its safety and effectiveness [13,29]. Therefore, effective communication campaigns should be designed to explain the safety of COVID-19 vaccines to the public and clarify misinformation or rumors, especially in HICs. Communication campaigns should be supported by the scientific community to address public concerns in COVID-19 vaccines. Through creating a space for a collaborative dialogue between the scientific community and the public, these campaigns would aim to not only provide the public with the latest information but also build public confidence in vaccine programs.

Social media users in HICs and LMICs expressed different concerns about vaccination convenience and accessibility. HIC users had less concerns about production and supply capacity of COVID-19 vaccines but more concerns on vaccine distribution and inequity. In contrast, LMIC users worried more about vaccine production and supply instead of vaccine distribution. Although HICs can ensure the supply of COVID-19 vaccines, systemic racial and economic disparities have existed for a long time in many fields, including health care and vaccination. The COVID-19 pandemic has disproportionately affected low-income groups and communities of color [24]. Disparities in access to COVID-19 vaccines possibly still occur and may be a point of concern that needs to be addressed in HICs. There is also much concern regarding access to COVID-19 vaccines and rollout gaps in LMICs compared with HICs. Many HICs have sought to gain priority access to COVID-19 vaccines by striking advance purchase agreements with vaccine manufacturers, instead of through WHO's global allocation mechanism, leaving few vaccine doses for LMICs [23,30]. Therefore, vaccine campaigns should be tailored to each context to address local concerns: LMICs should take efforts to address vaccine supply issues and HICs should focus more on equitable distribution within countries. This study calls for strengthened international partnerships and coordination to address the equitable access to COVID-19 vaccines worldwide, and the WHO should be empowered to take a leading role in guiding more preparedness actions to control the epidemic [31].

According to social media engagement analysis, we found that negative tweets had more followers and higher engagement than positive tweets. This finding might demonstrate that users on Twitter are more interested in communicating and disseminating negative messages such as those expressing misinformation or rumors and lack of vaccine confidence. Previous studies showed that Twitter users sharing misinformation tend to be more connected and clustered [32], and false information travels faster than true information does in social networks [33]. Although scientific experts received considerable attention early on in the COVID-19 pandemic, online attention shifted toward political communities as the pandemic developed [34]. During the evolution of the pandemic, scientific experts lost some of their influence, and it became harder for scientific information to reach a broad audience. In addition, the spread of misinformation or rumors on social media has been found to be significantly associated with vaccine hesitancy [35]. Social media is therefore a double-edged sword that not only can help disseminate public health knowledge directly to the public but also that can, through inappropriate use, be destructive to public health efforts, especially during a public health emergency [36-38]. Therefore, more efforts are needed to build a more proactive public health presence on social media, and health systems should *listen* to tweets from the public to help inform policies related to public health response.

Limitations

This study captured routine populations who may not be represented in traditional research designs, and social media data can eliminate reporting bias that occurs from speaking with a researcher [20]. However, this study has some limitations. First, there is an inherent bias shared among all social media studies, where users might present themselves differently online or represent a skewed-younger population [39]. Second, we did not extract demographic data because of difficulty to refer to users' profiles. Third, we used manual coding methods rather than automated annotation, which increased the length of time taken to annotate. Therefore, we only assessed vaccine hesitancy in large metropolises in a short period, and it reduced the generalizability of our findings. In the future, based on our coding data set from this study, we plan to develop a machine learning program to continuously track public attitudes toward COVID-19 vaccination.

Acknowledgments

The authors thank Xiaoyu Zhou, Qian Lv, and Simeng Hu from the School of Public Health, Fudan University, and Annika Wilder-Smith from Edinburgh Medical School for their help on data analysis. The authors also thank Fiona Yueqian Sun from the Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, for her help on Twitter data collection, and Xinyu Zhou from the School of Public Health, Fudan University for his help on Sina Weibo data collection.

ZH acknowledges financial support from the National Natural Science Foundation of China (No. 71874034) and the National Key R&D Program of China (No. 2018YFC1312600 and 2018YFC1312604). L Lin was supported by the UK Foreign, Commonwealth and Development Office and Wellcome (Grant no. 215373/Z/19/Z). The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

All data are publicly available.

Authors' Contributions

ZH and L Lin conceptualized and designed the study. ZH and SJP collected the data. YT, FD, L Lu, SZ, and KY analyzed the data. ZH and YT drafted the manuscript. ZH, L Lin, and HJL contributed toward the critical revision of the manuscript for important intellectual content. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

HJL is on the Merck Vaccine Confidence Advisory Board. Her research group, the Vaccine Confidence Project, received research grants from GSK and Merck on vaccine confidence issues. None of those research grants are related to this paper.

Multimedia Appendix 1
Coding framework for COVID-19 vaccine posts on social media.
[DOCX File , 20 KB - jmir_v23i6e27632_app1.docx]

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-12-23]
2. Coronavirus disease (COVID-19) advice for the public. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2020-12-23]
3. Chou WS, Budenz A. Considering emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. Health Commun 2020 Dec;35(14):1718-1722. [doi: 10.1080/10410236.2020.1838096] [Medline: 33124475]
4. Gao Q, Bao L, Mao H, Wang L, Xu K, Yang M, et al. Development of an inactivated vaccine candidate for SARS-CoV-2. Science 2020 Jul 03;369(6499):77-81 [FREE Full text] [doi: 10.1126/science.abc1932] [Medline: 32376603]
5. Haque A, Pant AB. Efforts at COVID-19 vaccined development: challenges and successes. Vaccines (Basel) 2020 Dec 06;8(4):739 [FREE Full text] [doi: 10.3390/vaccines8040739] [Medline: 33291245]
6. 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]
7. Bocquier A, Fressard L, Cortaredona S, Zaytseva A, Ward J, Gautier A, Baromètre santé 2016 group. Social differentiation of vaccine hesitancy among French parents and the mediating role of trust and commitment to health: A nationwide cross-sectional study. Vaccine 2018 Nov 29;36(50):7666-7673. [doi: 10.1016/j.vaccine.2018.10.085] [Medline: 30391054]
8. Larson HJ, Jarrett C, Schulz WS, Chaudhuri M, Zhou Y, Dube E, SAGE Working Group on Vaccine Hesitancy. Measuring vaccine hesitancy: the development of a survey tool. Vaccine 2015 Aug 14;33(34):4165-4175 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.037] [Medline: 25896384]

9. Chung Y, Schamel J, Fisher A, Frew PM. Influences on immunization decision-making among US parents of young children. Matern Child Health J 2017 Dec;21(12):2178-2187 [FREE Full text] [doi: 10.1007/s10995-017-2336-6] [Medline: 28755045]
10. Bianco A, Mascaro V, Zucco R, Pavia M. Parent perspectives on childhood vaccination: how to deal with vaccine hesitancy and refusal? Vaccine 2019 Feb 08;37(7):984-990. [doi: 10.1016/j.vaccine.2018.12.062] [Medline: 30655175]
11. Brown AL, Sperandio M, Turssi CP, Leite RMA, Berton VF, Succi RM, et al. Vaccine confidence and hesitancy in Brazil. Cad Saude Publica 2018 Sep 21;34(9):e00011618 [FREE Full text] [doi: 10.1590/0102-311X00011618] [Medline: 30281705]
12. Paterson P, Meurice F, Stanberry LR, Glismann S, Rosenthal SL, Larson HJ. Vaccine hesitancy and healthcare providers. Vaccine 2016 Dec 20;34(52):6700-6706 [FREE Full text] [doi: 10.1016/j.vaccine.2016.10.042] [Medline: 27810314]
13. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
14. Shah Z, Surian D, Dyda A, Coiera E, Mandl KD, Dunn AG. Automatically appraising the credibility of vaccine-related web pages shared on social media: a Twitter surveillance study. J Med Internet Res 2019 Nov 04;21(11):e14007 [FREE Full text] [doi: 10.2196/14007] [Medline: 31682571]
15. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
16. Steffens MS, Dunn AG, Wiley KE, Leask J. How organisations promoting vaccination respond to misinformation on social media: a qualitative investigation. BMC Public Health 2019 Oct 23;19(1):1348 [FREE Full text] [doi: 10.1186/s12889-019-7659-3] [Medline: 31640660]
17. Milinovich GJ, Williams GM, Clements ACA, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014 Feb;14(2):160-168 [FREE Full text] [doi: 10.1016/S1473-3099(13)70244-5] [Medline: 24290841]
18. Meltwater. URL: https://www.meltwater.com/ [accessed 2020-12-23]
19. COVID-19 vaccine tracker and landscape. World Health Organization. URL: https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines [accessed 2020-12-23]
20. Golder S, O'Connor K, Hennessy S, Gross R, Gonzalez-Hernandez G. Assessment of beliefs and attitudes about statins posted on Twitter: a qualitative study. JAMA Netw Open 2020 Jun 01;3(6):e208953 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8953] [Medline: 32584408]
21. 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]
22. Stone J, Can S. Linguistic analysis of municipal twitter feeds: factors influencing frequency and engagement. Government Inf Q 2020 Oct;37(4):101468 [FREE Full text] [doi: 10.1016/j.giq.2020.101468]
23. Wouters OJ, Shadlen KC, Salcher-Konrad M, Pollard AJ, Larson HJ, Teerawattananon Y, et al. Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment. Lancet 2021 Mar 13;397(10278):1023-1034 [FREE Full text] [doi: 10.1016/S0140-6736(21)00306-8] [Medline: 33587887]
24. Commissioners of the Lancet Commission on Vaccine Refusal, Acceptance, and Demand in the USA. Announcing the Lancet Commission on Vaccine Refusal, Acceptance, and Demand in the USA. Lancet 2021 Mar 27;397(10280):1165-1167. [doi: 10.1016/S0140-6736(21)00372-X] [Medline: 33639088]
25. Flaxman S, Mishra S, Gandy A, Unwin HJT, Coupland H, Mellan TA, Imperial College COVID-19 Response Team, et al. Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in European countries: technical description update. arXiv. Preprint posted online on April 23, 2020 [FREE Full text]
26. Pitzer VE, Chitwood M, Havumaki J, Menzies NA, Perniciaro S, Warren JL, et al. The impact of changes in diagnostic testing practices on estimates of COVID-19 transmission in the United States. Am J Epidemiol 2021 Apr 08:kwab089 [FREE Full text] [doi: 10.1093/aje/kwab089] [Medline: 33831148]
27. Gallup. Wellcome Global Monitor: how does the world feel about science and health? Libertatea. 2019. URL: https://static4.libertatea.ro/wp-content/uploads/2019/06/wellcome-global-monitor-2018.pdf [accessed 2019-06-01]
28. Kamidani S, Rostad CA, Anderson EJ. COVID-19 vaccine development: a pediatric perspective. Curr Opin Pediatr 2021 Feb 01;33(1):144-151. [doi: 10.1097/MOP.0000000000000978] [Medline: 33278108]
29. Mahase E. Covid-19: Oxford researchers halt vaccine trial while adverse reaction is investigated. BMJ 2020 Sep 09;370:m3525. [doi: 10.1136/bmj.m3525] [Medline: 32907856]
30. Lancet Commission on COVID-19 Vaccines and Therapeutics Task Force Members. Urgent needs of low-income and middle-income countries for COVID-19 vaccines and therapeutics. Lancet 2021 Feb 13;397(10274):562-564 [FREE Full text] [doi: 10.1016/S0140-6736(21)00242-7] [Medline: 33516284]
31. Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, et al. Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: infodemiology study. J Med Internet Res 2020 Aug 03;22(8):e21143 [FREE Full text] [doi: 10.2196/21143] [Medline: 32701460]
32. Pierri F, Piccardi C, Ceri S. Topology comparison of Twitter diffusion networks effectively reveals misleading information. Sci Rep 2020 Jan 28;10(1):1372. [doi: 10.1038/s41598-020-58166-5] [Medline: 31992754]

33. 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]
34. Durazzi F, Müller M, Salathé M, Remondini D. International expert communities on Twitter become more isolated during the COVID-19 pandemic. arXiv. Preprint posted online on November 11, 2020 [FREE Full text]
35. Aquino F, Donzelli G, De Franco E, Privitera G, Lopalco PL, Carducci A. The web and public confidence in MMR vaccination in Italy. Vaccine 2017 Aug 16;35(35 Pt B):4494-4498. [doi: 10.1016/j.vaccine.2017.07.029] [Medline: 28736200]
36. Pagoto S, Waring ME, Xu R. A call for a public health agenda for social media research. J Med Internet Res 2019 Dec 19;21(12):e16661 [FREE Full text] [doi: 10.2196/16661] [Medline: 31855185]
37. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
38. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct;5(10):e004206 [FREE Full text] [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]
39. Sloan L, Morgan J, Burnap P, Williams M. Who tweets? Deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. PLoS One 2015;10(3):e0115545 [FREE Full text] [doi: 10.1371/journal.pone.0115545] [Medline: 25729900]

Abbreviations

HIC: high-income country
LMIC: lower- and middle-income country
R&D: research and development
WHO: World Health Organization

Edited by C Basch; submitted 01.02.21; peer-reviewed by T Freeman, M Das; comments to author 14.04.21; revised version received 12.05.21; accepted 17.05.21; published 11.06.21.

Please cite as:
Hou Z, Tong Y, Du F, Lu L, Zhao S, Yu K, Piatek SJ, Larson HJ, Lin L
Assessing COVID-19 Vaccine Hesitancy, Confidence, and Public Engagement: A Global Social Listening Study
J Med Internet Res 2021;23(6):e27632
URL: https://www.jmir.org/2021/6/e27632
doi:10.2196/27632
PMID:34061757

Original Paper

Health Information Seeking Behaviors on Social Media During the COVID-19 Pandemic Among American Social Networking Site Users: Survey Study

Stephen Neely[1], PhD; Christina Eldredge[2], MD, PhD; Ron Sanders[3], DPA

[1]School of Public Affairs, College of Arts and Sciences, University of South Florida, Tampa, FL, United States
[2]School of Information, College of Arts and Sciences, University of South Florida, Tampa, FL, United States
[3]Florida Center for Cybersecurity, School of Information, College of Arts and Sciences, University of South Florida, Tampa, FL, United States

Corresponding Author:
Stephen Neely, PhD
School of Public Affairs
College of Arts and Sciences
University of South Florida
4202 E Fowler Ave
Tampa, FL, 33620
United States
Phone: 1 8139748423
Email: srneely@usf.edu

Related Articles:

Comment in: https://www.jmir.org/2022/2/e31415

Comment in: https://www.jmir.org/2022/2/e31569

Abstract

Background: In recent years, medical journals have emphasized the increasingly critical role that social media plays in the dissemination of public health information and disease prevention guidelines. However, platforms such as Facebook and Twitter continue to pose unique challenges for clinical health care providers and public health officials alike. In order to effectively communicate during public health emergencies, such as the COVID-19 pandemic, it is increasingly critical for health care providers and public health officials to understand how patients gather health-related information on the internet and adjudicate the merits of such information.

Objective: With that goal in mind, we conducted a survey of 1003 US-based adults to better understand how health consumers have used social media to learn and stay informed about the COVID-19 pandemic, the extent to which they have relied on credible scientific information sources, and how they have gone about fact-checking pandemic-related information.

Methods: A web-based survey was conducted with a sample that was purchased through an industry-leading market research provider. The results were reported with a 95% confidence level and a margin of error of 3. Participants included 1003 US-based adults (aged ≥18 years). Participants were selected via a stratified quota sampling approach to ensure that the sample was representative of the US population. Balanced quotas were determined (by region of the country) for gender, age, race, and ethnicity.

Results: The results showed a heavy reliance on social media during the COVID-19 pandemic; more than three-quarters of respondents (762/1003, 76%) reported that they have relied on social media at least “a little,” and 59.2% (594/1003) of respondents indicated that they read information about COVID-19 on social media at least once per week. According to the findings, most social media users (638/1003, 63.6%) were unlikely to fact-check what they see on the internet with a health professional, despite the high levels of mistrust in the accuracy of COVID-19–related information on social media. We also found a greater likelihood of undergoing vaccination among those following more credible scientific sources on social media during the pandemic (χ^2_{16}=50.790; ϕ=0.258; P<.001).

Conclusions: The findings suggest that health professionals will need to be both strategic and proactive when engaging with health consumers on social media if they hope to counteract the deleterious effects of misinformation and disinformation. Effective training, institutional support, and proactive collaboration can help health professionals adapt to the evolving patterns of health information seeking.

(J Med Internet Res 2021;23(6):e29802) doi:10.2196/29802

KEYWORDS

social media; internet; communication; public health; COVID-19; usage; United States; information seeking; web-based health information; survey; mistrust

Introduction

In recent years, medical journals have emphasized the increasingly critical role that social networking sites (SNSs) play in the dissemination of public health information and disease prevention guidelines [1,2]. Still, platforms such as Facebook and Twitter continue to pose unique challenges for clinical health care providers and public health officials alike. Although the public has grown more reliant on social media to stay informed during times of crisis [3], the information they receive comes from a variety of sources that are not always official or objective in nature. Health professionals often lack the time and resources that are necessary to keep pace with the rapidity of these web-based information environments. Moreover, effective health communications are increasingly complicated by factors such as politicization, antiscientific sentiments, and the potential that social networks have in rapidly spreading false information [4-6]. These challenges are perhaps the most acute under crisis conditions, which place unique strains on both health care systems and traditional information networks [7,8].

Although these concerns extend beyond the COVID-19 pandemic, their urgency has been underscored by the ongoing crisis. From the outset of the pandemic, public health officials noted an alarming volume of erroneous misinformation (as well as malicious disinformation) associated with COVID-19 on social media. Estimates from early studies have suggested that as much as 25% of the COVID-19–related information circulating on platforms like Twitter may contain some degree of misinformation [9]. The World Health Organization labeled this phenomenon as an “infodemic” and suggested that such misinformation can “lead to poor observance of public health measures, thus reducing their effectiveness and endangering countries’ ability to stop the pandemic” [10].

In order to effectively communicate during public health emergencies, it is increasingly critical for health professionals to understand how patients gather health-related information on the internet and adjudicate the merits of such information. However, while much has been written about social media’s expanding role in health communications, very little empirical data have been collected to examine how the public actually uses social media to learn and stay informed about ongoing health emergencies. To that end, we conducted a survey of 1003 American adults in order to better understand how heavily they have relied on social media and the specific ways in which they have used social media to learn about the COVID-19 pandemic. In light of the growing concerns over the proliferation of misinformation on social media, this study also aims to aid both health care practitioners and researchers in understanding how SNS users interact with and rely on credible scientific sources and how they have gone about fact-checking pandemic-related information. Collectively, this study seeks to better inform health communications through an enriched understanding of social media’s evolving role in health information seeking.

Methods

A web-based survey of 1003 US-based adults was conducted (January 9 to January 12, 2021) through Prodege MR, an industry-leading market research provider. This survey was funded by Florida’s Center for Cybersecurity at the University of South Florida. Survey respondents were selected by using a stratified quota sampling approach to ensure that the sample was representative. Balanced quotas were determined (by region of the country) for gender, age, race, ethnicity, and education based on the US Census Bureau’s 2019 American Community Survey (ACS). Table 1 provides a summary of the comparison between sample respondents and the 2019 ACS data.

The initial sampling target was a total of 1067 individuals (which represents a margin of error 3.0 for the US population). However, data cleaning revealed a small number of incomplete and unusable responses, resulting in a total sample size of 1003 (a margin of error 3.09 for the US population). The results were reported with a 95% confidence level and a margin of error of 3.09. It is worth noting that the data collection method necessarily excluded those who lack internet access. However, given that the focus of the study was on social media users, this did not represent a serious threat to validity. Perhaps more importantly, the method naturally underrepresented those with lower levels of education. We made deliberate attempts to target this group during survey administration, though a gap persisted for those lacking a high school diploma (Table 1).

Table 1. Sample comparison.[a]

Characteristic	Sample in this study, %	2019 American Community Survey, %
Gender		
Female	51.2	51.1
Male	48.8	48.9
Age (years)		
18-24	11.3	11.9
25-34	17.8	17.8
35-44	17	16.5
45-54	16	16
55-64	17.1	16.6
≥65	20.8	21.2
Race		
Black or African American	15.7	15.1
White	73.2	76.4
Asian or Pacific Islander	7.9	7.6
American Indian or Alaska Native	1	0.8
Other	2.2	0.2
Ethnicity		
Hispanic	17.7	17.5
Non-Hispanic	82.3	82.5
Education		
Less than high school	6.8	11.4
High school or equivalent	25.9	27.6
Some college or an associate degree	34	30.4
4-year degree	21.5	19.3
Graduate or professional degree	11.7	11.4

[a] Data are from the Florida Center for Cybersecurity’s January 2021 COVID-19 survey.

Results

Summary of Results

Although the growing importance of social media in health communications has been widely discussed, our understanding of these trends, particularly those at the consumer level, has been largely anecdotal. The results reported below add some empirical context to our general understanding of these trends while also contributing important new data to enhance our understanding of how SNS users seek out and fact-check medical information during a public health emergency. The results are reported in subsections that follow the organization of the survey.

Reliance and Confidence

The results from the survey affirm that Americans have relied heavily on social media to stay informed about COVID-19. Among the survey respondents, more than three-quarters (762/1003, 76%) stated that they have relied on social media at least “a little” to stay informed about the COVID-19 pandemic, while just under half (458/1003, 45.6%) reported that they have relied on social media “a lot” (Table 2). Further, 59.2% (594/1003) of respondents indicated that they read information about COVID-19 on social media at least once per week, while roughly one-third (323/1003, 32.2%) do so every day. These responses highlight the extent to which social media has become a primary source of health information for Americans; a large number of users (762/1003, 76%) reported that they do not merely encounter such content on the internet but also rely on platforms such as Facebook and Twitter for health information.

Table 2. Reliance on social media for pandemic-related information.[a]

Questions and responses	Respondents, n (%)
How much have you relied on social media to stay informed about the COVID-19 pandemic?	
A great deal	208 (20.7)
A lot	250 (24.9)
A little	304 (30.3)
Not at all	241 (24)
On average, how often do you read information about COVID-19 on social media?	
Every day	323 (32.2)
A few days a week	271 (27)
Once a week	112 (11.2)
Less often	297 (29.6)
I am confident in the accuracy of the information I see about COVID-19 on social media	
Strongly agree	73 (7.3)
Somewhat agree	250 (24.9)
Neither agree nor disagree	252 (25.1)
Somewhat disagree	188 (18.7)
Strongly disagree	240 (23.9)

[a] Data are from the Florida Center for Cybersecurity's January 2021 COVID-19 survey.

Although people's reliance on SNSs has been remarkably high throughout the pandemic, only about one-third of respondents (323/1003, 32.2%) expressed confidence in the accuracy of the information that they receive about COVID-19 on social media. Although paradoxical, this finding is consistent with prior studies that have shown a similar lack of confidence in the accuracy of news and political information on the internet, despite the increased reliance on such web-based media for information seeking [11]. In our study, this may, at least in part, reflect the sharp politicization of the COVID-19 pandemic [12]. Among the survey respondents, more three-quarters (761/1003, 76.1%) agreed that "politics has made it harder to learn the truth about Covid-19." For health professionals, this highlights the extent to which the politicization of public health efforts can obscure scientific guidance and complicate health communications, especially when treatment and mitigation become matters of public policy.

Fact-checking Social Media

It has been recently been noted in medical literature that identifying and mitigating misinformation on social media is a growing priority for health professionals [4]. It has also been pointed out that doing so will require proactive steps, such as "meeting people where they are and through the information networks and devices they use for day-to-day interactions" [13]. With these goals in mind, it is important for health professionals to understand the propensity of social media users to validate and verify (ie, fact-check) potential misinformation that they encounter on the internet and to understand the types of sources that social media users are turning to for guidance during public health crises. Prior research has suggested that rigorous fact-checking is less common on social media [14], and the survey responses appear to affirm this suggestion. Only about one-third of respondents (365/1003, 36.4%) indicated that they have "talked to a doctor or healthcare professional about the accuracy of something they saw on social media related to Covid-19" (Table 3). In contrast, respondents were almost twice as likely to have "talked to friends, family, or coworkers" about such information (686/1003, 68.4%), and 22% of respondents were more likely to have conducted their own internet research (578/1003, 57.6%) to fact-check COVID-19–related information on social media.

Table 3. Fact-checking pandemic-related information on social media.[a]

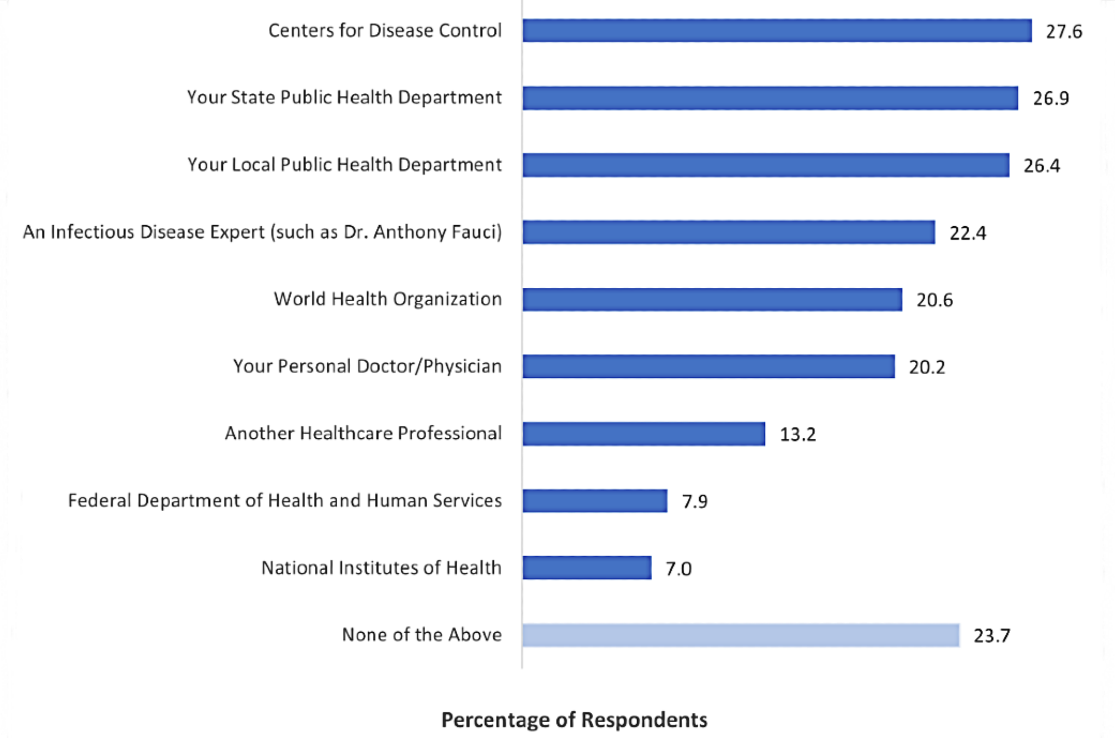
Responses to the following question: "Please indicate whether or not you've done each of the following since the start of the pandemic?"	Respondents, n (%)
"…done internet research to fact-check something that I saw on social media related to COVID-19"	578 (57.6)
"…talked to friends, family, or coworkers about the accuracy of something I saw on social media related to COVID-19"	686 (68.4)
"…talked to my doctor or a healthcare professional about the accuracy of something that I saw on social media related to COVID-19"	365 (36.4)

[a] Data are from the Florida Center for Cybersecurity's January 2021 COVID-19 survey.

Following Scientific Sources

Although only about one-third of respondents reported actively fact-checking information with a medical professional, 581/762 (76.2%) reported that they followed at least 1 authoritative scientific source on social media during the pandemic (Figure 1). The data in Figure 1 only represent those who reported at least “a little” reliance on social media to stay informed about COVID-19 (n=762). More than one-quarter of respondents began following the Centers for Disease Control (210/762, 27.6%), their state public health department (205/762, 26.9%), or their local public health department (201/762, 26.4%). Just under one-quarter of respondents reported following an infectious disease expert such as Dr Anthony Fauci (171/762, 22.4%). Further, one-fifth of respondents (154/762, 20.2%) reported following their own personal doctor or physician, while 101/762 (13.2%) began following “another healthcare professional.”

Figure 1. Health Information Sources "Followed" (As % of Respondents).



Perhaps surprisingly, the data showed that respondents were slightly more likely to have followed institutional actors on social media (ie, the Centers for Disease Control and Prevention or a public health department) than they were to have followed individual health experts (ie, their personal physician or an infectious disease expert). This is somewhat inconsistent with findings from prior research, which have suggested that individual actors are typically more influential on social media [15,16]. The nature and severity of the COVID-19 pandemic may account for this difference, though additional research into this topic is warranted in order to better inform professional best practices.

On one hand, the responses were promising in the sense that most social media users did appear to have intentionally expanded their web-based networks during the pandemic to include credible institutional and individual medical and scientific sources (ie, 581/762 [76.2%] have followed at least 1 authoritative scientific source). On the other hand, it was noteworthy that nearly one-quarter of those who relied on social media for pandemic-related information (181/762, 23.8%) did not choose to follow such sources, and only one-fifth of respondents (154/762, 20.2%) began engaging with their own personal physician (ie, the health care provider whom they are most likely to trust for personalized health guidance) on social media.

Scientific Credibility and Vaccine Intentions

As previous literature has suggested, the dissemination of credible scientific information on social media is necessary for disease prevention and effective public health management. The crosstabs from our survey highlighted one specific example of this; those who opted to follow credible scientific sources were significantly more likely to indicate that they will accept a COVID-19 vaccine ($P<.001$). When asked, 588/1003 (58.6%) of respondents to the survey reported that they would definitely or probably undergo vaccination [17]. Those who followed at least 2 credible public health or medical sources were 10% more likely to indicate that they would “definitely get vaccinated” than those who did not follow any such sources (Table 4). Those who followed 4 or more such sources were over 25% more likely to report the same response.

Table 4. Medical sources and vaccine intentions.[a]

Intentions	Number of medical sources followed				
	None	1	2-3	4-5	≥6
Will definitely undergo vaccination, %	28.2	31.7	38.8	55	55.6
Will probably undergo vaccination, %	24.9	23	24.9	17.5	14.8
May or may not undergo vaccination, %	16.6	15.5	19.6	23.8	18.5
Will probably not undergo vaccination, %	14.4	11.7	6.2	0	3.7
Will definitely not undergo vaccination, %	16	18.1	10.5	3.8	7.4

[a] Chi-square test results: χ^2_{16}=50.790; ϕ=0.258; P<.001.

In contrast, nearly one-third (55/181, 30.4%) of those who did not follow any public health or medical sources said that they would either "probably not" or "definitely not" undergo vaccination. This number fell by roughly 50% among those following at least 2 of the aforementioned sources. A chi-square test showed that the differences were statistically significant (χ^2_{16}=50.790; ϕ=0.258; P<.001). This observed relationship likely reflects some degree of simultaneity. However, the data did suggest that exposure to credible scientific information on the internet is positively related to compliance with pandemic mitigation measures. Given the high vaccination levels needed to achieve herd immunity, closing even small informational gaps could prove critical to ending the COVID-19 pandemic.

Discussion

Data from the Pew Research Center show that social media is slowly but steadily supplanting traditional information mediums as a primary source of news and information for many Americans [18]. It has become increasingly clear from existing literature (including our own research) that this trend includes personal and public health information seeking behaviors as well. In this study, we surveyed 1003 American adults on their use of social media to learn about the COVID-19 pandemic. The survey responses confirmed that health consumers have relied heavily on social media to stay up to date with and informed about the COVID-19 pandemic. More than three-quarters (762/1003, 76%) of respondents stated that they have relied on social media at least "a little," and 59.2% (594/1003) of respondents indicated that they read information about COVID-19 on social media at least once per week. The heavy reliance on social media observed among US-based SNS users is consistent with findings from recent research conducted in various international settings, including China [19] and Europe [20].

Our findings also showed that only about one-third of SNS users (365/1003, 36.4%) have fact-checked pandemic-related information with a medical professional, despite the widespread distrust in the accuracy of COVID-19–related information that is shared on social media. Although disconcerting, this observation is consistent with the previous finding that rigorous fact-checking is relatively uncommon on social media [14]. We also found a greater likelihood of undergoing vaccination among those following more credible scientific sources on social media (χ^2_{16}=50.790; ϕ=0.258; P<.001), suggesting that scientific credibility may be crucial when promoting compliance with public health guidelines. Recent research has suggested that social media has been instrumental in the spread of vaccine hesitancy [21,22], thereby underscoring the need for health professionals and scientific experts to actively engage with patients and health consumers on social networks in order to address common misconceptions about vaccine safety and efficacy.

Our findings highlight the increasing importance of social media in health information seeking and thus highlight its potential value to health professionals as a conduit for personal and public health communications. However, the growing popularity of SNS platforms for health information seeking is not without its potential drawbacks. Among such drawbacks is the noted propensity for SNSs to facilitate the rapid and widespread dissemination of misinformation and disinformation [4,5,23]. Several studies have examined the presence and effects of misinformation related to COVID-19 since the start of the pandemic. One early analysis of pandemic-related messaging on Twitter suggested that as much as 25% of COVID-19–related information that is being circulated on the platform may contain some degree of misinformation [9]. Recent studies have found that exposure to misinformation on the internet is linked to decreases in the awareness of and compliance with preventative and mitigation measures [24,25]. Although the anonymous and instantaneous nature of social networks can contribute to the rapid spread of health-related misinformation, some research has suggested that social media may also offer an effective avenue for health professionals to counter speculation and misinformation. For example, in a recent experimental study, corrective infographics circulated by the World Health Organization were found to reduce scientific misperceptions about COVID-19 prevention [26].

Another potential drawback of people's increasing reliance on SNS platforms is the potential for social media overload to increase anxiety and adversely impact the psychological well-being of patients and SNS users. Several recent studies have documented this propensity during the COVID-19 pandemic, though these concerns are likely to be germane to any sustained public health or emergency scenario. One study of young SNS users in the United Kingdom found that information overload related to the COVID-19 pandemic resulted in diminished psychological well-being, including unhealthy levels of the fear of COVID-19 [27]. A similar study that was conducted in Hong Kong during 2020 found a correlation between social media usage and pandemic-related anxiety as well as diminished social trust in information [28]. Although our results showed an increased reliance on social media for health information seeking, these previous findings have suggested that this trend may have adverse mental health impacts for some SNS users—a fact that health professionals will need to be increasingly cognizant of when considering best practices for health communications.

When put into context with the emerging body of literature, our findings suggest that health professionals will need to become increasingly savvy when it comes to social media usage—not just reactively (ie, “setting the record straight”) but also proactively. Given the fact that 76% (762/1003) of health consumers in this survey have relied on social media at least “a little” as a source of health information during the pandemic, accurate and consistent messaging by credible public health organizations is just a start. Based on our research, we believe that this will require more active engagement between health professionals and patients and consumers. To achieve this degree of engagement, health professionals and public health organizations will need to cultivate and customize state-of-the-art social engineering skills to include data mining and natural language processing skills as well as skills that can only be called “active measures” (ie, those for monitoring, anticipating, and responding to misinformation and disinformation on social media platforms), especially during a public health crisis [29]. Furthermore, future research should explore recommendations for institutional policies regarding social media use by government and public health institutions [30].

At the patient level, given the apparent reticence of many social media users to connect with health professionals when fact-checking web-based information, it may be necessary for providers to more deliberately engage patients in conversations about the medical information that they are encountering on social media. Along with raising these issues individually in clinical settings, health care providers can also leverage the evolving functionality of platforms such as Twitter and Facebook to organize live question-and-answer sessions or fact-checking sessions for their patients and communities. For example, during the Zika virus outbreak, the US Department of Health and Human Services held digital town halls via social media. These were routinely advertised through posts such as the following:

> *Don't miss the @HHSGov #AtoZika Twitter Town Hall tomorrow, Aug. 30, 10 a.m. Submit questions using #AtoZika. #Zika* [31]

More deliberate networking efforts between local providers and public health agencies may be an effective means of organizing and promoting such events in order to counter misinformation on the internet.

Although these types of broadly targeted communications are critical to effective pandemic management, recent research has also suggested that for many SNS users, personal appeals from reputable practicing physicians may be a more effective means of public health messaging via social media [32]. This type of messaging may be of particular importance in the case of politicized public health emergencies, such as the COVID-19 pandemic. For example, one recent study found that reliance and trust in institutional information sources declined between March and April of 2020 as the pandemic became increasingly politicized [33]. However, our findings suggest that even during the COVID-19 pandemic, a notably low number of Americans have chosen to follow their own clinical providers on social media. Although this may be due to personal choice, the lack of a health care provider presence on social media could be a factor of that choice. For example, a prior survey of physicians indicated that confidentiality, organizational support, and time are all significant barriers to social media adoption [30]. Physicians, physician assistants, and nurse practitioners may need more training on how to effectively use social media to engage with and inform their patient populations. Among other things, such training should address health care providers’ concerns regarding governance, ethics, trust, and patient privacy [34,35]. Notably, prominent organizations such as the Mayo Clinic have already begun incorporating social media literacy into the academic training of health professionals [36].

It is important to acknowledge that adding social media–related duties to the duties of an already overextended health care workforce may further exacerbate the burnout that is experienced by many care providers [37]. However, these data underscore the increasing tendency of patients and consumers to rely on SNS platforms for health-related information. We believe that effective training, institutional support, and proactive collaboration can help health professionals adapt to the evolving patterns of health information seeking behaviors while also protecting the well-being of providers, especially in the midst of an already taxing pandemic.

Acknowledgments

Funding for this project was provided by the Florida Center for Cybersecurity at the University of South Florida.

Conflicts of Interest

None declared.

References

1. Merchant RM. Evaluating the potential role of social media in preventive health care. JAMA 2020 Feb 04;323(5):411-412. [doi: 10.1001/jama.2019.21084] [Medline: 31922532]

2. Merchant RM, Elmer S, Lurie N. Integrating social media into emergency-preparedness efforts. N Engl J Med 2011 Jul 28;365(4):289-291. [doi: 10.1056/NEJMp1103591] [Medline: 21793742]
3. Lachlan KA, Spence PR, Lin X, Najarian K, Del Greco M. Social media and crisis management: CERC, search strategies, and Twitter content. Comput Human Behav 2016 Jan;54:647-652. [doi: 10.1016/j.chb.2015.05.027]
4. Chou WYS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
5. Conrado SP, Neville K, Woodworth S, O'Riordan S. Managing social media uncertainty to support the decision making process during emergencies. J Decis Syst 2016 Jun 16;25(sup1):171-181 [FREE Full text] [doi: 10.1080/12460125.2016.1187396]
6. Hughes AL, Palen L. The evolving role of the public information officer: An examination of social media in emergency management. J Homel Secur Emerg Manag 2012;9(1):1-20. [doi: 10.1515/1547-7355.1976]
7. Pechta LE, Brandenburg DC, Seeger MW. Understanding the dynamics of emergency communication: Proposition for a four-channel model. J Homel Secur Emerg Manag 2010;7(1):1-18. [doi: 10.2202/1547-7355.1671]
8. Steelman TA, Nowell B, Bayoumi D, McCaffrey S. Understanding information exchange during disaster response: Methodological insights from infocentric analysis. Adm Soc 2012 Dec 23;46(6):707-743. [doi: 10.1177/0095399712469198]
9. Kouzy R, Jaoude JA, 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]
10. Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. World Health Organization. 2020 Sep 23. URL: https://tinyurl.com/k4zrca9e [accessed 2021-04-01]
11. Shearer E, Matsa KE. News use across social media platforms 2018. Pew Research Center. 2018 Sep 10. URL: https://www.journalism.org/wp-content/uploads/sites/8/2018/09/PJ_2018.09.10_social-media-news_FINAL.pdf [accessed 2021-03-25]
12. Wood S, Schulman K. Beyond politics - Promoting Covid-19 vaccination in the United States. N Engl J Med 2021 Feb 18;384(7):e23. [doi: 10.1056/NEJMms2033790] [Medline: 33406324]
13. Merchant RM, South EC, Lurie N. Public health messaging in an Era of social media. JAMA 2021 Jan 19;325(3):223-224. [doi: 10.1001/jama.2020.24514] [Medline: 33393964]
14. Jun Y, Meng R, Johar GV. Perceived social presence reduces fact-checking. Proc Natl Acad Sci U S A 2017 Jun 06;114(23):5976-5981 [FREE Full text] [doi: 10.1073/pnas.1700175114] [Medline: 28533396]
15. Hagen L, Keller T, Neely S, DePaula N, Robert-Cooperman C. Crisis communications in the age of social media. Soc Sci Comput Rev 2017 Aug 21;36(5):523-541. [doi: 10.1177/0894439317721985]
16. Whelan E, Golden W, Donnellan B. Digitising the R&D social network: revisiting the technological gatekeeper. Information Systems Journal 2011 Sep 14;23(3):197-218. [doi: 10.1111/j.1365-2575.2011.00384.x]
17. Taking America's Temperature on the Vaccine: COVID-19 Vaccine & Policy Survey. CyberFlorida. 2021 Jan 14. URL: https://cyberflorida.org/report/vaccine-and-policy/ [accessed 2021-03-25]
18. Shearer E. Social media outpaces print newspapers in the US as a news source. Pew Research Center. 2018 Dec 10. URL: https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/ [accessed 2021-04-10]
19. 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]
20. Stjernswärd S, Ivert AK, Glasdam S. Perceptions and effects of COVID-19 related information in Denmark and Sweden - a web-based survey about COVID-19 and social media. Z Gesundh Wiss 2021 Apr 26:1-15. [doi: 10.1007/s10389-021-01539-5] [Medline: 33936931]
21. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct;5(10):e004206 [FREE Full text] [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]
22. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593. [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
23. Venegas-Vera AV, Colbert GB, Lerma EV. Positive and negative impact of social media in the COVID-19 era. Rev Cardiovasc Med 2020 Dec 30;21(4):561-564 [FREE Full text] [doi: 10.31083/j.rcm.2020.04.195] [Medline: 33388000]
24. Lee JJ, Kang KA, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and Preventive Behaviors: Cross-sectional online study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
25. Reyes LM, Ortiz L, Abedi M, Luciano Y, Ramos W, Reyes PJDJ. Misinformation on COVID-19 origin and its relationship with perception and knowledge about social distancing: A cross-sectional study. PLoS One 2021 Mar 09;16(3):e0248160. [doi: 10.1371/journal.pone.0248160] [Medline: 33690685]
26. Vraga EK, Bode L. Addressing COVID-19 misinformation on social media preemptively and responsively. Emerg Infect Dis 2021 Feb;27(2):396-403 [FREE Full text] [doi: 10.3201/eid2702.203139] [Medline: 33395379]

27. Liu H, Liu W, Yoganathan V, Osburg VS. COVID-19 information overload and generation Z's social media discontinuance intention during the pandemic lockdown. Technol Forecast Soc Change 2021 May;166:120600. [doi: 10.1016/j.techfore.2021.120600]
28. Wong FHC, Liu T, Leung DKY, Zhang AY, Au WSH, Kwok WW, et al. Consuming information related to COVID-19 on social media among older adults and its association with anxiety, social trust in information, and COVID-safe behaviors: Cross-sectional telephone survey. J Med Internet Res 2021 Feb 11;23(2):e26570 [FREE Full text] [doi: 10.2196/26570] [Medline: 33523825]
29. Islam MS, Sarkar T, Khan SH, Kamal AHM, Hasan SMM, Kabir A, et al. COVID-19-related infodemic and its impact on public health: A global social media analysis. Am J Trop Med Hyg 2020 Oct;103(4):1621-1629 [FREE Full text] [doi: 10.4269/ajtmh.20-0812] [Medline: 32783794]
30. Grajales 3rd FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
31. @HHSGov. Save the date! We're hosting an #AtoZika Town Hall on Tues (8/30) starting at 10am ET. Submit Qs using #AtoZika. Twitter. 2016 Aug 28. URL: https://twitter.com/hhsgov/status/770017877797826560 [accessed 2021-03-01]
32. Solnick RE, Chao G, Ross RD, Kraft-Todd GT, Kocher KE. Emergency physicians and personal narratives improve the perceived effectiveness of COVID-19 public health recommendations on social media: A randomized experiment. Acad Emerg Med 2021 Feb;28(2):172-183 [FREE Full text] [doi: 10.1111/acem.14188] [Medline: 33263357]
33. Ali SH, Foreman J, Tozan Y, Capasso A, Jones AM, DiClemente RJ. Trends and predictors of COVID-19 information sources and their relationship with knowledge and beliefs related to the pandemic: Nationwide cross-sectional study. JMIR Public Health Surveill 2020 Oct 08;6(4):e21071 [FREE Full text] [doi: 10.2196/21071] [Medline: 32936775]
34. Panahi S, Watson J, Partridge H. Social media and physicians: Exploring the benefits and challenges. Health Informatics J 2016 Jun;22(2):99-112 [FREE Full text] [doi: 10.1177/1460458214540907] [Medline: 25038200]
35. Farnan JM, Sulmasy LS, Worster BK, Chaudhry HJ, Rhyne JA, Arora VM, American College of Physicians Ethics, Professionalism and Human Rights Committee, American College of Physicians Council of Associates, Federation of State Medical Boards Special Committee on Ethics and Professionalism. Online medical professionalism: patient and public relationships: policy statement from the American College of Physicians and the Federation of State Medical Boards. Ann Intern Med 2013 Apr 16;158(8):620-627 [FREE Full text] [doi: 10.7326/0003-4819-158-8-201304160-00100] [Medline: 23579867]
36. Cabrera D. Mayo clinic includes social media scholarship activities in academic advancement. Mayo Clinic Social Media Network. 2016 May 25. URL: https://socialmedia.mayoclinic.org/2016/05/25/mayo-clinic-includes-social-media-scholarship-activities-in-academic-advancement/ [accessed 2021-02-28]
37. Lacy BE, Chan JL. Physician burnout: The hidden health care crisis. Clin Gastroenterol Hepatol 2018 Mar;16(3):311-317. [doi: 10.1016/j.cgh.2017.06.043] [Medline: 28669661]

Abbreviations

ACS: American Community Survey
SNS: social networking site

Edited by C Basch; submitted 20.04.21; peer-reviewed by M Amini, G Kernohan; comments to author 22.05.21; revised version received 26.05.21; accepted 26.05.21; published 11.06.21.

Please cite as:
Neely S, Eldredge C, Sanders R
Health Information Seeking Behaviors on Social Media During the COVID-19 Pandemic Among American Social Networking Site Users: Survey Study
J Med Internet Res 2021;23(6):e29802
URL: https://www.jmir.org/2021/6/e29802
doi:10.2196/29802
PMID:34043526

Original Paper

Real-time Prediction of the Daily Incidence of COVID-19 in 215 Countries and Territories Using Machine Learning: Model Development and Validation

Yuanyuan Peng[1*], BSc; Cuilian Li[2*], MBBS; Yibiao Rong[3], PhD; Chi Pui Pang[2,4], DPhil; Xinjian Chen[1], PhD; Haoyu Chen[2], MD

[1]School of Electronics and Information Engineering, Soochow University, Suzhou, China
[2]Joint Shantou International Eye Center, Shantou University & the Chinese University of Hong Kong, Shantou, China
[3]College of Engineering, Shantou University, Shantou, China
[4]Department of Ophthalmology and Visual Sciences, Chinese University of Hong Kong, Hong Kong, China (Hong Kong)
[*]these authors contributed equally

Corresponding Author:
Haoyu Chen, MD
Joint Shantou International Eye Center
Shantou University & the Chinese University of Hong Kong
Joint Shantou International Eye Center
North Dongxia Road
Shantou, 515041
China
Phone: 86 075488393560
Email: drchenhaoyu@gmail.com

Abstract

Background: Advanced prediction of the daily incidence of COVID-19 can aid policy making on the prevention of disease spread, which can profoundly affect people's livelihood. In previous studies, predictions were investigated for single or several countries and territories.

Objective: We aimed to develop models that can be applied for real-time prediction of COVID-19 activity in all individual countries and territories worldwide.

Methods: Data of the previous daily incidence and infoveillance data (search volume data via Google Trends) from 215 individual countries and territories were collected. A random forest regression algorithm was used to train models to predict the daily new confirmed cases 7 days ahead. Several methods were used to optimize the models, including clustering the countries and territories, selecting features according to the importance scores, performing multiple-step forecasting, and upgrading the models at regular intervals. The performance of the models was assessed using the mean absolute error (MAE), root mean square error (RMSE), Pearson correlation coefficient, and Spearman correlation coefficient.

Results: Our models can accurately predict the daily new confirmed cases of COVID-19 in most countries and territories. Of the 215 countries and territories under study, 198 (92.1%) had MAEs <10 and 187 (87.0%) had Pearson correlation coefficients >0.8. For the 215 countries and territories, the mean MAE was 5.42 (range 0.26-15.32), the mean RMSE was 9.27 (range 1.81-24.40), the mean Pearson correlation coefficient was 0.89 (range 0.08-0.99), and the mean Spearman correlation coefficient was 0.84 (range 0.2-1.00).

Conclusions: By integrating previous incidence and Google Trends data, our machine learning algorithm was able to predict the incidence of COVID-19 in most individual countries and territories accurately 7 days ahead.

(J Med Internet Res 2021;23(6):e24285) doi:10.2196/24285

KEYWORDS

COVID-19; daily incidence; real-time prediction; machine learning; Google Trends; infoveillance; infodemiology; digital health; digital public health; surveillance; prediction; incidence; policy; prevention; model

Introduction

COVID-19, a highly infectious disease with serious clinical manifestations, was first reported in China in late 2019 and spread to other countries within weeks [1,2]. It was announced as a public health emergency of international concern by the World Health Organization (WHO) on January 30, 2020, and declared a pandemic on March 11, 2020 [3]. As of August 21, 2020, COVID-19 has been reported in 215 countries and territories, with 22,876,009 confirmed cases and 797,289 deaths [4]. The severity of the pandemic is variable in individual countries and territories. In some regions, severe outbreaks, occurred and case numbers took a long time to decrease. In other regions, it took more time for the spread of the disease to affect people. Even in places where the spread has been brought under control, there have been second or third waves of outbreaks, with variable severities of morbidities and mortalities.

Prediction of the incidence of COVID-19 in individual countries and territories is extremely important to provide reference for governments, health care providers, and the general public to prepare management measures to combat the disease. In the early months of 2020, prediction was useful to inform the countries and territories at risk of outbreak to take action for prevention. Currently, although the COVID-19 outbreak has already occurred in almost all regions globally, prediction has merits in monitoring the severity of spreading and recovery and assessing the likelihood of a secondary or tertiary epidemic. COVID-19 may become a seasonal or persistent epidemic in future years, like influenza. Thus, predicting and monitoring COVID-19 activity is needed at present and in the future. As a typical example, Google search data can predict the incidence of influenza effectively [5].

There are reported studies on the prediction of the incidence of COVID-19. A susceptible-exposed-infectious-recovered metapopulation model was used to simulate the epidemic in major cities in China [1]. Our previous study showed that data from internet search engines and social media platforms could predict the incidence of COVID-19 in China [6]. Later, data mining and machine learning techniques were used to predict the incidence of COVID-19 using data from the previous incidence or internet search volumes [7,8]. In these studies, only a few countries were investigated or the world was considered as one region. Furthermore, as the pandemic continues, increasing amounts of data are available for modeling and calibration. The actual situation may be different from previous predictions. For example, Tuli et al [9] predicted that the total number of new cases would reach 97% of the total expected cases on August 14 in the United States; however, actually, there were still tens of thousands of new cases daily. Therefore, an up-to-date and evolving prediction method is needed.

In this study, we aim to develop an efficient and novel methodology for real-time prediction of COVID-19 activity based on the previous daily incidence of COVID-19 and infoveillance data (search volume data via Google Trends) in all individual countries and territories worldwide.

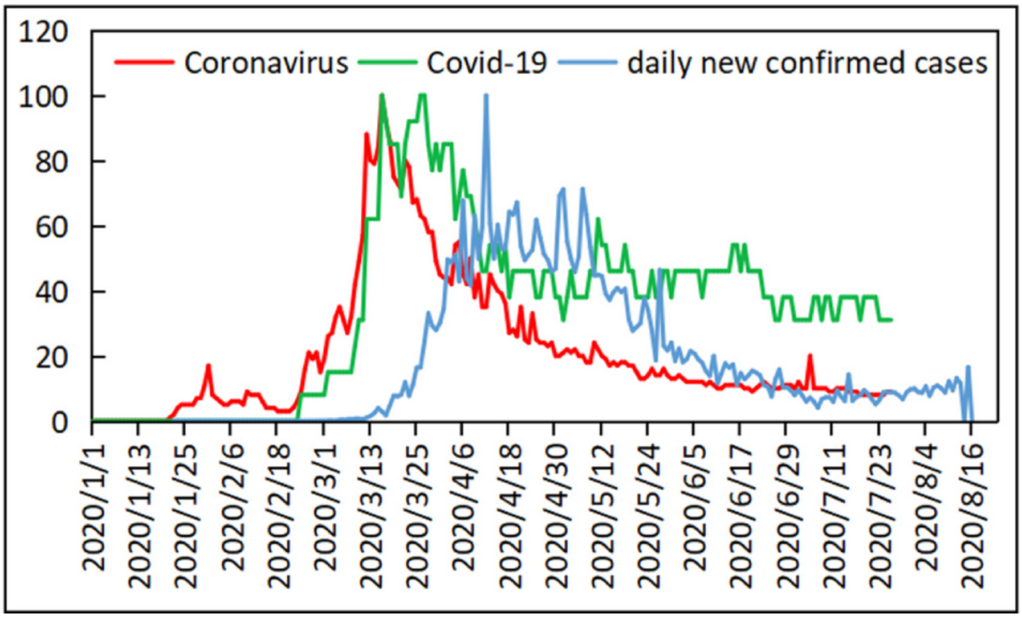
Methods

Data Source

Two sets of data were collected. The first data set was the search volume data obtained from the Google Trends service. We collected the Google search volumes of 28 candidate features related to COVID-19 from January 1 to July 26, 2020, in 215 countries and territories. We used 14 terms, including *coronavirus*, *pneumonia*, and *Covid-19*; 6 symptom-related terms [10] (*cough*, *diarrhea*, *fatigue*, *fever*, *nasal congestion*, and *rhinorrhea*); and 5 prevention-related terms (*hand washing*, *hand sanitizer*, *mask*, *social distance*, and *social isolation*. There were two main reasons for selecting these 14 terms for Google Trends data collection: (1) Previous studies [6,10-13] have shown that the internet search data for these terms were correlated with the incidence of COVID-19. (2) These 14 terms are related to COVID-19 in different aspects, such as topics, symptoms, and prevention; therefore, they can reflect its incidence. Two types of Google Trends data were retrieved. The first type was the data of interest over time, defined as the search interest relative to the highest point for the specific term, region, and time interval. Values were calculated on a scale from 0 to 100, where 100 is the peak popularity for the time period. The second type of data, data of interest by region, was retrieved by setting the region to “worldwide” for the given term and time. The values were calculated on a scale from 0 to 100, where 100 is the location with the most popularity as a fraction of total searches in that location. It is notable that we included low search volume regions to obtain data for more regions.

The second data set was the daily number of new COVID-19 cases from January 10 to August 16, 2020, in 215 countries and territories, obtained from the WHO website [14]. For consistency, the values of the daily number of new COVID-19 cases in the 215 countries and territories were normalized to a scale from 0 to 100. An example showing the search volumes of the terms *coronavirus* and *Covid-19* and the daily incidence of COVID-19 in the United Kingdom is shown in Figure 1.

Figure 1. The Google Trends search volumes for the topics *coronavirus* (red line) and *Covid-19* (green line) and the number of daily new confirmed cases (blue line) in the United Kingdom. All values are scaled from 1 to 100.



Correlation Analysis and Feature Engineering

We conducted two types of correlation analyses in each country or territory to find the right combination of input features. The first analysis was the Spearman correlation between the Google volume data of each feature and daily new confirmed cases with n days of lag (n=7, 14, 21, 28). We obtained the average and maximum Spearman correlation coefficients for different lag days between each Google volume data feature and daily newly confirmed cases in the 215 countries and territories (Multimedia Appendix 1). Each Google volume data feature was related to the daily new confirmed cases in the 215 countries and territories. The top related item was the interest over time in *coronavirus*. We found a better correlation with the 21-day lag than with the other days of lag. Therefore, we combined 28 Google volume data features with daily new confirmed cases of 21-day lag and formed 29 input features.

We calculated the Spearman correlation among 28 independent Google Trends volume data features in the 215 countries and territories to assess the independence of the features. We obtained the average results of the Spearman correlation coefficients between the top 10 cross-correlation Google volume data features in the 215 countries and territories. The highest correlation coefficient, between *coronavirus* interest over time and interest by region, was 0.59, which is lower than the criterion of strong correlation (0.8). Additionally, the cross-correlation between the remaining 18 Google volume data features was less than 0.1. Therefore, the correlation analysis results showed that the 28 Google volume data features were independent.

Cluster Analysis

The internet search patterns and the management of COVID-19 vary among countries and territories. We used a hierarchical clustering technique to cluster the 215 countries and territories into several groups, and we built a model for each group. Two types of data were used for clustering. One type was the related metrics of the importance scores of 29 input features resulting from the random forest regression algorithm in 215 countries and territories. The other type was the correlation of the development trend of daily new confirmed cases in the 215 countries and territories. Finally, the 215 countries and territories were clustered into 8 groups (Multimedia Appendix 2). There were 150 countries and territories in the first group, 38 in the second group, 11 in the third group, 7 in the fourth group, 6 in the fifth group, and 1 country or territory each in the sixth, seventh, and eighth groups.

In addition, to explore the internal relationships between the countries and territories in each cluster, two types of Spearman correlation coefficients were calculated. The first was the Spearman correlation between the daily new confirmed cases among the countries in the first five groups, which contain more than one country or territory. The second was the Spearman correlation between the search volume of the term *coronavirus* among the countries in the first five groups. These two correlations were also calculated in the sixth, seventh, and eighth groups with a single country or territory. The average Spearman correlation coefficients are shown in Table 1. There were two findings. The first was a correlation between the incidence of COVID-19 among the countries and territories in each group, especially in the first group, which contains 150 countries and territories. The second was that the correlation of Google search data among countries and territories in the first five groups was high. These findings also showed the reliability of our clustering method.

Table 1. Average Spearman correlation coefficients in each group of countries/territories for the 2 data sets.

Data set	Average Spearman correlation coefficient							
	Group 1 (n=150)	Group 2 (n=38)	Group 3 (n=11)	Group 4 (n=7)	Group 5 (n=6)	Group 6 (n=1)	Group 7 (n=1)	Group 8 (n=1)
Daily new cases	0.46	0.18	0.35	0.26	0.38	0.23	0.04	0.15
Coronavirus search volume	0.76	0.68	0.71	0.74	0.69	0.52	0.00	0.43

Modeling and Evaluation

We built a model for each group of countries/territories separately according to the clustering results and produced a 7-day-ahead incidence prediction of COVID-19 for the 215 countries and territories. The random forest regression algorithm has many decision trees; therefore, it has good robustness and a strong ability to resist overfitting. In addition, the random forest regression (RFR) algorithm gave the importance score of the features, which was helpful for feature selection. Therefore, the RFR algorithm was used to forecast the daily new confirmed cases of COVID-19 over the time series data set, which consisted of Google Trends data and the incidence of COVID-19 with 21-day lag provided by the WHO. Python 3.7.6 was used for the modeling and evaluation.

To quantitatively evaluate the performance of the models, we calculated four metrics: the mean absolute error (MAE), root mean square error (RMSE), Pearson correlation coefficient, and Spearman correlation coefficient.

where p_i, t_i, and M represent the predicted value, true value, and total number of days; X and Y are two independent time series; and σ_X and σ_Y are the standard deviations of X and Y, respectively. x_i and y_i are the ith value of X and Y and and are the mean values of X and Y, respectively, and n is the length of the time series. The P values of the Pearson correlations were calculated to assess the statistical significance.

Optimization and Validation

Several experiments were conducted to optimize and validate the results. First, to evaluate the role of each input feature in the prediction, we calculated the importance score of each input feature. To validate the features included in this study, a series of ablation studies, in which the top n features (n=1, 2, ..., 29) were used as inputs in turn, were conducted on different groups of countries/territories according to the clustering results. The Pearson correlation coefficient was averaged for the 8 groups of countries/territories based on the sum of the input features (Figure 2). According to the results, the top 13, 10, 10, 25, 2, 5, 1, and 3 features were used for the respective 8 groups in further experiments. The models using previous daily incidence with and without Google Trends data as input features were compared (Table 2). Adding Google search volume features as input improved the MAE, RMSE, Pearson correlation, and Spearman correlation by 0.28, 0.88, 0.06, and 0.08, respectively, in terms of average results in 215 countries and territories. These results confirmed that the Google Trends data played a positive role in predicting the incidence of COVID-19.

Second, to reduce the influence of random noise, we used multiple-step forecasting (5-step, 6-step, 7-step, 8-step, 9-step, and 10-step), which used the data over the past n days (n=5, 6, 7, 8, 9, 10) to predict the daily new confirmed cases of COVID-19. The average quantitative results of the different steps of the second period in the 215 countries and territories are shown in Table 3. The performance of the 7-step prediction was slightly better than that obtained using other numbers of steps in terms of 4 of 5 quantitative metrics in our study. Therefore, we selected the 7-step prediction.

Third, the proposed RFR algorithm was compared with two other mainstream algorithms. One was decision tree regression (DTR), which is a traditional machine learning algorithm. The other was long short-term memory (LSTM), which is a deep learning algorithm. Especially, LSTM, an artificial recurrent neural network, was used as a 3-layer model in our study. TensorFlow and Keras were used as frameworks for training the LSTM models. Table 4 shows the average quantitative results of the different methods for the 215 countries and territories. The random forest regression algorithm achieved better performance in terms of all quantitative metrics compared to DTR and LSTM.

Fourth, we trained the models with as much data as possible. Therefore, we updated the models and repeated the experiments at regular intervals. The effectiveness of the models at 3 time periods was compared.

Fifth, we predicted the incidence of COVID-19 2 days and 7 days ahead. Table 5 shows the average quantitative results of different time windows of the second period in the 215 countries and territories. We found the predictive power of the 7-day window to be less accurate than that of the 2-day window. It was noted that the 2-day window prediction may not be sufficiently informative for policy making, and the 7-day window would be more useful to enable a responsible authority to respond proactively and impose interventions in time. Therefore, we used the 7-day window for the prediction of the incidence of COVID-19.

Figure 2. The average Pearson correlation coefficients of the number of top-scoring features included in the prediction of the incidence of COVID-19 in different groups of countries. The red, blue, green, orange, purple, yellow, black, and brown lines represent groups 1 to 8, respectively.

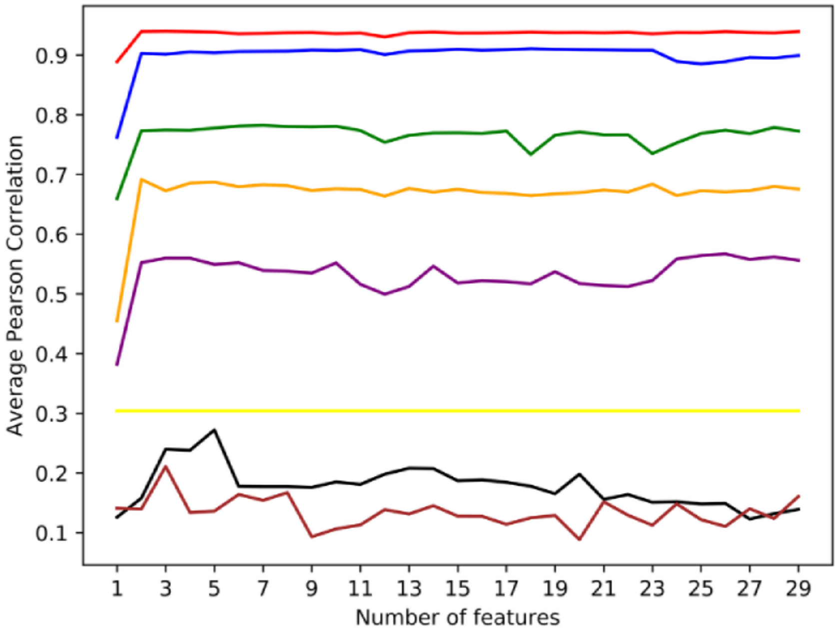


Table 2. Average results without and with Google Trends data in 215 countries.

Method	MAE[a]			RMSE[b]			Pearson correlation			Spearman correlation			*P* value[c]		
	Avg[d]	Max[e]	Min[f]	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min
Without Google Trends	5.67	17.20	0.77	10.27	26.00	2.86	0.82	0.99	–0.02	0.76	0.99	–0.02	.018	<.999	<.001
With Google Trends	5.42	15.32	0.26	9.27	24.40	1.81	0.89	0.99	0.08	0.84	1.00	0.21	.001	.24	<.001

[a] MAE: mean absolute error.

[b] RMSE: root mean square error.

[c] *P* value of the Pearson correlation.

[d] Avg: average.

[e] Max: maximum.

[f] Min: minimum.

Table 3. The results of multiple-step forecasting in 215 countries and territories in the second period.

Method	MAE[a]	RMSE[b]	Pearson correlation	Spearman correlation	*P* value[c]
Step_5	5.48	9.35	0.88	0.84	.006
Step_6	5.46	9.21	0.88	0.84	.006
Step_7	5.44	9.12	0.88	0.84	.006
Step_8	5.66	9.49	0.88	0.84	.006
Step_9	5.66	9.46	0.88	0.85	.006
Step_10	5.70	9.50	0.88	0.84	.006

[a] MAE: mean absolute error.

[b] RMSE: root mean square error.

[c] *P* value of the Pearson correlation.

Table 4. Average performance of different methods in 215 countries and territories.

Method	MAE[a]	RMSE[b]	Pearson correlation	Spearman correlation	*P* value[c]
Decision tree regression	6.78	11.57	0.81	0.79	.011
Long short-term memory	9.13	14.29	0.76	0.78	.025
Random forest regression	5.42	9.27	0.89	0.84	.006

[a] MAE: mean absolute error.

[b] RMSE: root mean square error.

[c] *P* value of the Pearson correlation.

Table 5. The average results of different time windows of the second period in 215 countries and territories.

Time window	MAE[a]	RMSE[b]	Pearson correlation	Spearman correlation	*P* value[c]
2 days ahead	4.09	7.40	0.94	0.87	.003
7 days ahead	5.42	9.27	0.89	0.84	.006

[a] MAE: mean absolute error.

[b] RMSE: root mean square error.

[c] *P* value of the Pearson correlation.

Results

We produced 7-day-ahead and real-time COVID-19 forecasts for 215 countries and territories. Our final models performed well in predicting the daily new confirmed cases of COVID-19 in most of the countries and territories examined. Of the 215 countries and territories, 198 (92.1%) had MAEs <10 (Figure 3a), and 187 (87.0%) had Pearson correlation coefficients >0.8 (Figure 3b). In all 215 countries/territories, the mean MAE was 5.42 (range 0.26-15.32), the mean RMSE was 9.27 (range 1.81-24.40), the mean Pearson correlation coefficient 0.89 (range 0.08-0.99) and mean Spearman correlation coefficient 0.84 (range 0.21-1.00), with *P*<.001 in most of the countries and territories. Detailed results of the individual countries and territories are listed in Multimedia Appendix 3. The average results for each group are shown in Table 6. Generally, the models performed well in the first, second, third, and fifth groups, which contained 150, 38, 11, and 6 countries and territories, respectively. The Pearson correlation coefficients were >0.80 in most of these countries and territories. In contrast, the performance was poorer in groups 4, 6, 7, and 8. The Pearson correlation coefficients were <0.80 in most of these countries.

The performance of the models in 3 different time periods is shown in Table 7. As the training data increased, the overall prediction performance of the model improved. In addition, we randomly selected 4 countries from the 150 countries in the first group for further analysis. The predicted and actual daily new confirmed cases of the 3 different time periods in these 4 countries are shown in Figure 4. For example, our models can predict the increase in daily new confirmed cases in China from July 18 to July 31 and the decrease of daily new confirmed cases in China from August 3 to August 16. Our models were able to predict the daily new confirmed case numbers consistently.

Figure 3. Heat maps of the (a) mean absolute error and (b) Pearson correlation coefficients of the predicted and actual daily new confirmed case numbers in different countries worldwide. The deeper the color, the lower the mean absolute error (a) and the higher the correlation coefficient (b).

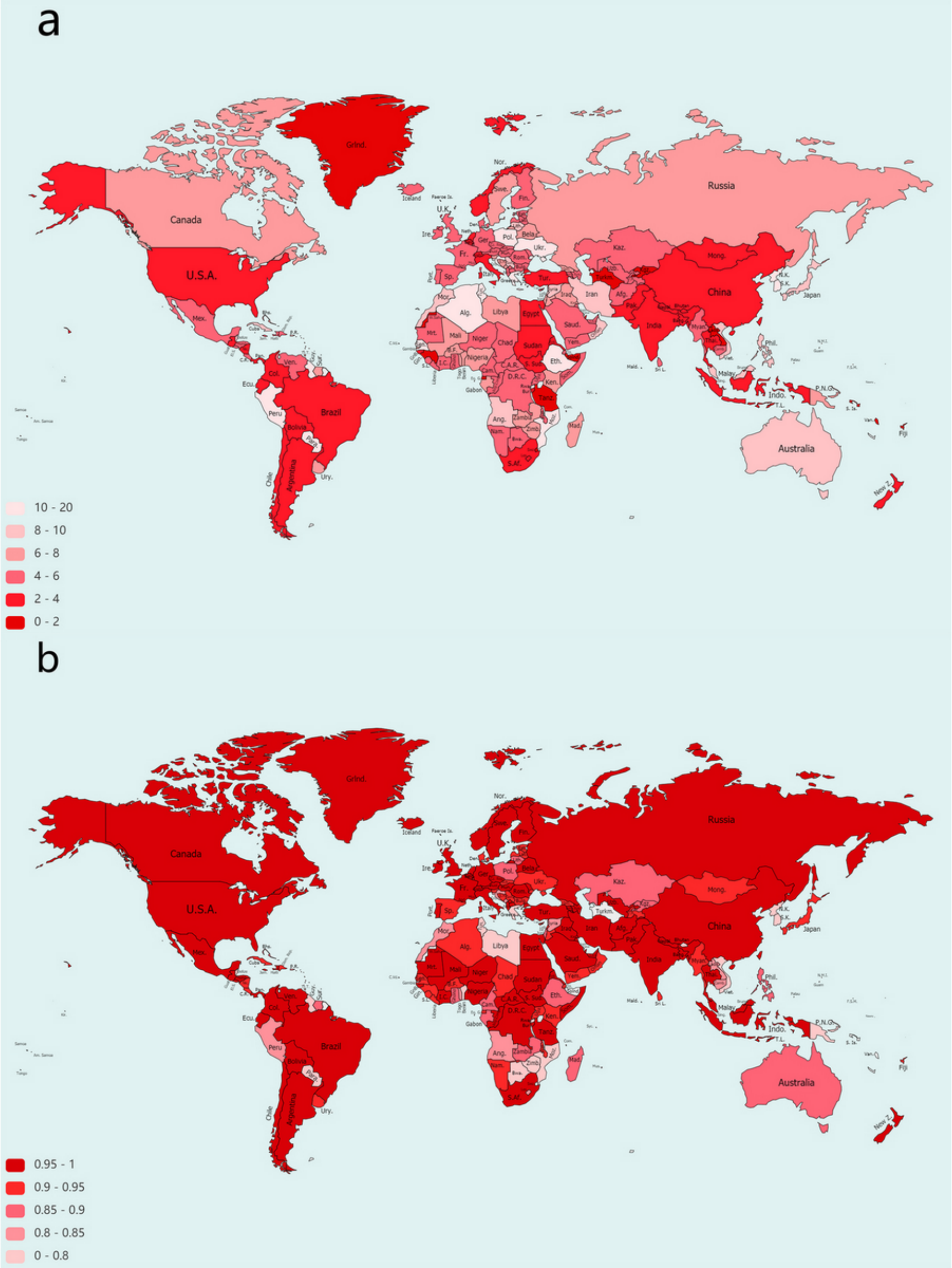


Table 6. Performance of the final models in the different groups of countries/territories.

Cluster	Countries/territories, n	MAE[a]			RMSE[b]			Pearson correlation			Spearman correlation			*P* value[c]		
		Avg[d]	Max[e]	Min[f]	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min
1	150	5.52	15.00	0.26	8.96	23.45	3.32	0.94	0.99	0.72	0.92	1.00	0.74	<.001	<.001	<.001
2	38	3.92	10.18	0.36	7.17	14.73	1.81	0.90	0.98	0.65	0.67	0.90	0.25	<.001	<.001	<.001
3	11	7.38	13.26	1.71	13.77	20.68	5.97	0.77	0.89	0.70	0.81	0.92	0.63	<.001	<.001	<.001
4	7	6.41	8.82	1.32	14.01	16.86	8.83	0.52	0.63	0.39	0.60	0.76	0.42	<.001	<.001	<.001
5	6	6.12	15.32	1.76	13.01	23.19	4.99	0.66	0.86	0.50	0.71	0.79	0.58	<.001	<.001	<.001
6	1	5.72	5.72	5.72	12.75	12.75	12.75	0.27	0.27	0.27	0.58	0.58	0.58	<.001	<.001	<.001
7	1	2.39	2.39	2.39	2.39	12.17	12.17	0.30	0.30	0.30	0.46	0.46	0.46	<.001	<.001	<.001
8	1	14.60	14.60	14.60	24.40	24.40	24.40	0.008	0.008	0.008	0.21	0.21	0.21	.24	.24	.24
Total	215	5.42	15.32	0.26	9.27	24.40	1.81	0.89	0.99	0.08	0.84	1.00	0.21	.001	.24	<.001

[a] MAE: mean absolute error.

[b] RMSE: root mean square error.

[c] *P* value of the Pearson correlation.

[d] Avg: average.

[e] Max: maximum.

[f] Min: minimum.

Table 7. Performance of the models in different time periods.

Time period (2020)	MAE[a]			RMSE[b]			Pearson correlation			Spearman correlation			*P* value[c]		
	Avg[d]	Max[e]	Min[f]	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min
Feb 04-July 17	10.11	20.88	0.31	15.76	33.15	3.94	0.79	0.99	-0.03	0.75	1.00	0.04	.009	<.999	<.001
Feb 04-July 31	5.44	15.98	0.28	9.12	25.02	2.71	0.88	0.99	-0.05	0.84	1.00	-0.13	.006	.83	<.001
Feb 04-Aug 16	5.42	15.32	0.26	9.27	24.20	1.81	0.89	0.99	0.08	0.84	1.00	0.21	.001	.24	<.001

[a] MAE: mean absolute error.

[b] RMSE: root mean square error.

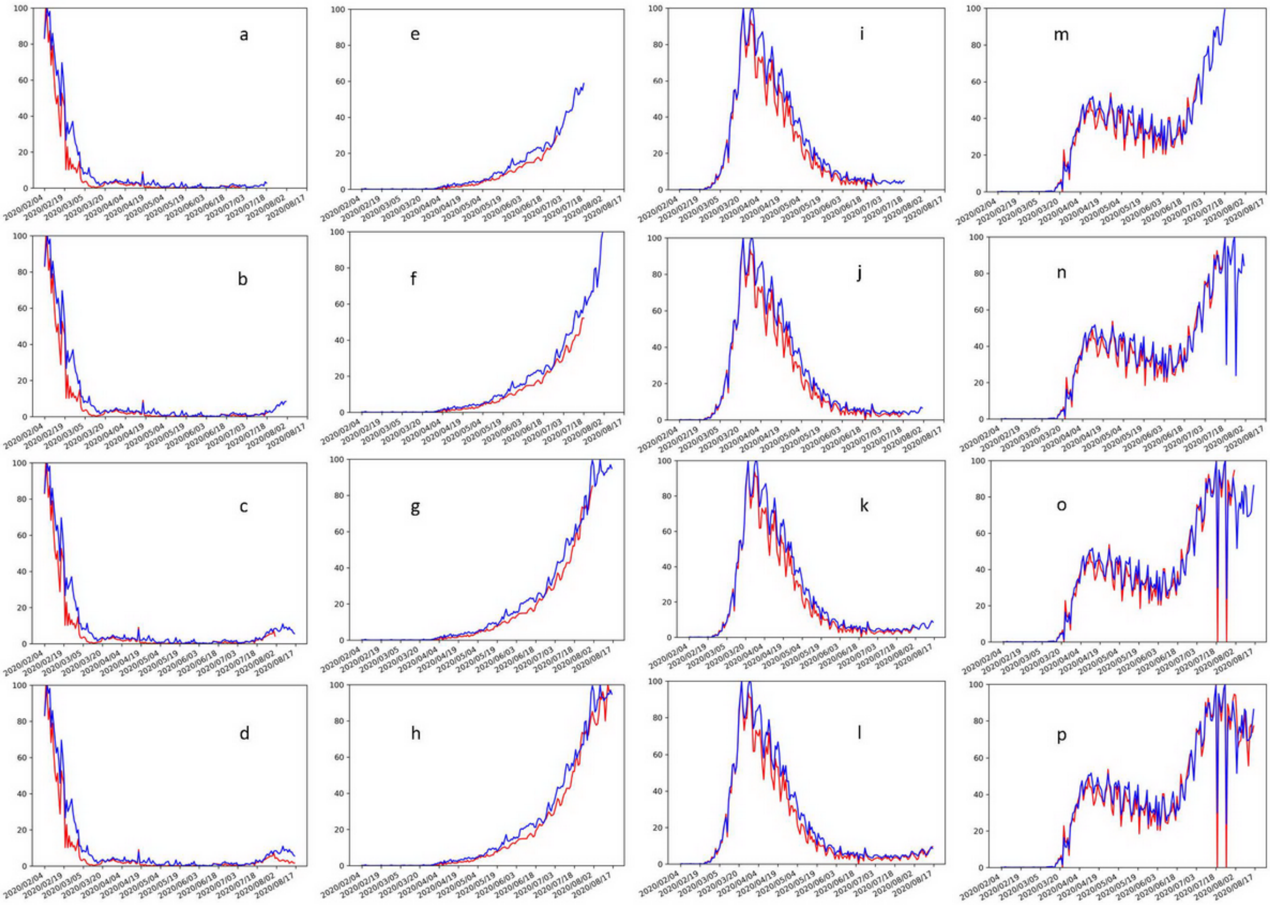
[c] *P* value of the Pearson correlation.

[d] Avg: average.

[e] Max: maximum.

[f] Min: minimum.

Figure 4. The prediction curves in China (a-d), India (e-h), Italy (i-l), and the United States (m-p). The blue and red lines indicate the predicted and real daily new cases, respectively. The first panels in each country (a, e, i, m) show the prediction for July 3 to July 17. The second panels (b, f, j, n) show the actual number of daily new cases for July 3 to July 17 and the prediction for July 18 to July 31. The third panels (c, g, k, o) show the actual number of daily new cases for July 18 to July 31 and the prediction for August 1 to August 16. The fourth panels (d, h, l, p) show the actual number of daily new cases from August 1 to August 16.



Discussion

Principal Results

In this study, we have established a method to obtain a 7-day-ahead prediction of COVID-19 activity by combining the previous incidence of COVID-19 and Google search volume data at the country or territory level in the real world. The models performed well in most countries. In a total of 215 countries and territories, the mean MAE, RMSE, Pearson correlation and Spearman correlation values were 5.42, 9.27, 0.89, and 0.878, respectively. The *P* value was <.001 in most countries and territories, which shows the significance of the Pearson correlation coefficient. The results show that the longer the training time and the greater the amount of available data, the higher the prediction accuracy.

Comparison With Prior Work

Our study has several advantages compared to other reported studies on similar topics. First, we investigated 215 individual countries and territories. Other studies only investigated a single country [15] or several countries [16], or they considered the world as one whole region [11,17]. Our method performed well in most of the countries and territories. Second, we used clustering methods to pool data with similar patterns together. The advantage of large amounts of data improved the performance of the models. Our results show that the groups with a large number of countries and territories had better results compared to those with a limited number of countries. There has been a report on clustering of provinces in China [12], but not clustering of countries. In another study, countries were grouped into 4 clusters; however, this clustering was not used for prediction [18]. Third, we combined data regarding both previous incidence of COVID-19 and Google Trends search volumes for prediction. Both of these data sources have been used individually in forecasting COVID-19 incidence. Our results showed that adding the Google Trends data improved the performance of the models developed from the previous incidence data. Fourth, we used machine learning methods for modeling. Many previous studies [13,19,20] indicated that Google Trends data are correlated with the incidence of COVID-19. Machine learning has been used to manage some large data sets and develop models to predict the incidence of COVID-19 in only a few countries [20]. Our models can produce useful and reliable predictions in most countries. Some studies used epidemical models such as susceptible-infected-recovered-dead (SIRD) for prediction. However, the spread of COVID-19 is affected by many factors, including public health management, traveling, and other social activities, which are not taken into account in the SIRD model. In Iran, COVID-19 was expected to subside at the beginning of May 2020 according to the SIRD model. However, at the time of writing this manuscript, thousands of new cases were still being reported daily in Iran [21]. Fifth, our models used the most up-to-date data and were updated three times. As more data become available, the performance of the models can be further improved. Some models developed several months ago may have severe errors. For example, Tuli et al [9] predicted that the total number of new cases would reach 97% of the total expected cases on August 14, 2020, in the United States [9]; however, tens of thousands of new cases were reported daily in July and August.

The high accuracy of our models enables real-time forecasting of the short-term trends of COVID-19, not only during the outbreak but also during the recovery period and subsequent second or third epidemic in individual countries. The methods can also be adapted to predict the incidence of subregions. Real-time digital surveillance of COVID-19 is provided, which would save time and resources in data collection. As the COVID-19 pandemic is changing rapidly, digital health systems should provide an effective solution to address the challenges to public health and consequential socioeconomical complications with high efficiency. The epidemic of COVID-19 is likely to persist in the future, and it may even become a seasonal infectious disease like influenza. Therefore, accurate surveillance and prediction of its activity would help governments, health care providers and the general population to take appropriate actions to compact the disruptive effects of COVID-19 [5].

Limitations

We recognize some limitations of the current study. First, the models did not predict well in some countries, which were clustered into independent groups and had limited periods of data for training. Second, the occurrence of COVID-19 in some countries would have increased the awareness of other countries, especially those with a close relationship, consequently resulting in a large Google search volume for terms related to COVID-19. Third, there is a limitation in applying our models in countries and territories where Google is not the mainstream search engine. Thus, predictions of COVID-19 incidence in regions with limited internet access or prohibited Google access may not be accurate. Fourth, COVID-19 and influenza share some common symptoms and even prevention methods. Therefore, the Google Trends data of these terms may not be able to differentiate COVID-19 and influenza. Further studies are needed to develop an algorithm to differentiate these two infectious diseases using Google Trends.

Conclusions

In this study, we integrated the Google Trends data and previous daily incidence of 215 individual countries and territories using the techniques of features engineering, country clustering, and machine learning. We are able to achieve prediction 7 days ahead of time of the daily incidence of COVID-19 in real time in most countries and territories.

Acknowledgments

This study was supported in part by the National Key R&D Program of China (2018YFA0701700), the Department of Education of Guangdong Province (2020KZDZX1086), the National Natural Science Foundation of China (61971298), and the Grant for Key Disciplinary Project of Clinical Medicine under the Guangdong High-level University Development Program.

Authors' Contributions

YP and YR analyzed the data. YP drafted the manuscript. CL collected the data. CL, CPP, HC, and XC interpreted the data. HC and XC designed the study. CP and HC revised the manuscript. All authors agreed on the final version for submission to the journal.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Spearman correlation coefficients of 28 Google volume data features with the incidence of COVID-19 at n (n=7, 14, 21, 28) days of lag in 215 countries.

[DOCX File , 19 KB - jmir_v23i6e24285_app1.docx]

Multimedia Appendix 2

List of countries and territories with different clustering results.

[DOCX File , 16 KB - jmir_v23i6e24285_app2.docx]

Multimedia Appendix 3

Performance of the predicting model in 215 individual countries and territories.

[DOCX File , 56 KB - jmir_v23i6e24285_app3.docx]

References

1. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb 29;395(10225):689-697 [FREE Full text] [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]
2. Del Rio C, Malani PN. 2019 novel coronavirus-important information for clinicians. JAMA 2020 Mar 17;323(11):1039-1040. [doi: 10.1001/jama.2020.1490] [Medline: 32022836]
3. Heymann DL, Shindo N, WHO Scientific and Technical Advisory Group for Infectious Hazards. COVID-19: what is next for public health? Lancet 2020 Feb 22;395(10224):542-545 [FREE Full text] [doi: 10.1016/S0140-6736(20)30374-3] [Medline: 32061313]
4. COVID-19 coronavirus pandemic. Worldometer. URL: https://www.worldometers.info/coronavirus/#countries [accessed 2021-06-10]
5. 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]
6. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10) [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
7. Chakraborty T, Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: a data-driven analysis. Chaos Solitons Fractals 2020 Jun;135:109850 [FREE Full text] [doi: 10.1016/j.chaos.2020.109850] [Medline: 32355424]
8. Yang Z, Zeng Z, Wang K, Wong S, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis 2020 Mar;12(3):165-174 [FREE Full text] [doi: 10.21037/jtd.2020.02.64] [Medline: 32274081]
9. Tuli S, Tuli S, Tuli R, Gill S. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet Things 2020 Sep;11:100222 [FREE Full text] [doi: 10.1016/j.iot.2020.100222]
10. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
11. Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY, Snot Force Alliance. Correlations of online search engine trends with coronavirus disease (COVID-19) incidence: infodemiology study. JMIR Public Health Surveill 2020 May 21;6(2):e19702 [FREE Full text] [doi: 10.2196/19702] [Medline: 32401211]
12. Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis JT, et al. A machine learning methodology for real-time forecasting of the 2019-2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models. ArXiv. Preprint posted online on April 08, 2020. [Medline: 32550248]
13. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: data mining and deep learning pilot study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
14. Coronavirus disease (COVID-19) situation reports. World Health Organization. 2020 Feb 13. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-11-17]
15. Ribeiro MHDM, da Silva RG, Mariani VC, Coelho LDS. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos Solitons Fractals 2020 Jun;135:109853 [FREE Full text] [doi: 10.1016/j.chaos.2020.109853] [Medline: 32501370]
16. Stübinger J, Schneider L. Epidemiology of coronavirus COVID-19: forecasting the future incidence in different countries. Healthcare (Basel) 2020 Apr 15;8(2) [FREE Full text] [doi: 10.3390/healthcare8020099] [Medline: 32326512]
17. Mavragani A. Tracking COVID-19 in Europe: infodemiology approach. JMIR Public Health Surveill 2020 Apr 20;6(2):e18941 [FREE Full text] [doi: 10.2196/18941] [Medline: 32250957]
18. Khan MHR, Hossain A. Countries are clustered but number of tests is not vital to predict global COVID-19 confirmed cases: a machine learning approach. medRxiv. Preprint posted online on April 2, 2020. [doi: 10.1101/2020.04.24.20078238]
19. Noori A, Mikaeil R, Mokhtarian M, Haghshenas S, Foroughi M. Feasibility of intelligent models for prediction of utilization factor of TBM. Geotech Geol Eng 2020 Jan 28;38(3):3125-3143 [FREE Full text] [doi: 10.1007/s10706-020-01213-9]
20. Yuan X, Xu J, Hussain S, Wang H, Gao N, Zhang L. Trends and prediction in daily new cases and deaths of COVID-19 in the United States: an internet search-interest based model. Explor Res Hypothesis Med 2020 Apr 18;5(2):1-6 [FREE Full text] [doi: 10.14218/ERHM.2020.00023] [Medline: 32348380]
21. Ndiaye BM, Tendeng L, Seck D. Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting. ArXiv. Preprint posted online on April 3, 2020 [FREE Full text]

Abbreviations

DTR: decision tree regression
LSTM: long short-term memory
MAE: mean absolute error

RFR: random forest regression
RMSE: root mean square error
SIRD: susceptible-infected-recovered-dead
WHO: World Health Organization

Edited by C Basch; submitted 13.09.20; peer-reviewed by A Al-Hasan, R Mpofu; comments to author 06.11.20; revised version received 03.04.21; accepted 31.05.21; published 14.06.21.

Please cite as:
Peng Y, Li C, Rong Y, Pang CP, Chen X, Chen H
Real-time Prediction of the Daily Incidence of COVID-19 in 215 Countries and Territories Using Machine Learning: Model Development and Validation
J Med Internet Res 2021;23(6):e24285
URL: https://www.jmir.org/2021/6/e24285
doi:10.2196/24285
PMID:34081607

Original Paper

Estimating the COVID-19 Spread Through Real-time Population Mobility Patterns: Surveillance in Low- and Middle-Income Countries

Stefanos Tyrovolas[1,2,3,4], PhD; Iago Giné-Vázquez[2,3], PhD; Daniel Fernández[5,6], PhD; Marianthi Morena[4], MSc; Ai Koyanagi[3,7], MD, PhD; Mark Janko[8], PhD; Josep Maria Haro[2,3], MD, PhD; Yang Lin[1], PhD; Paul Lee[1,9], PhD; William Pan[8], DrPH; Demosthenes Panagiotakos[4], PhD; Alex Molassiotis[1], PhD

[1]School of Nursing, Hong Kong Polytechnic University, Kowloon, Hong Kong
[2]Parc Sanitari Sant Joan de Déu, Universitat de Barcelona, Fundació Sant Joan de Déu, Barcelona, Spain
[3]Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental, Madrid, Spain
[4]Nutrition and Dietetics Department, School of Health Science and Education, Harokopio University, Athens, Greece
[5]Department of Statistics and Operations Research, Polytechnic University of Catalonia (UPC)-BarcelonaTech, Barcelona, Spain
[6]Institute of Mathematics of Polytechnic University of Catalonia (UPC)-BarcelonaTech (IMTech), Barcelona, Spain
[7]Catalan Institution for Research and Advanced Studies, Barcelona, Spain
[8]Duke University, Durham, NC, United States
[9]Department of Health Sciences, University of Leicester, Leicester, United Kingdom

Corresponding Author:
Stefanos Tyrovolas, PhD
School of Nursing
Hong Kong Polytechnic University
PQ402
Hung Hom
Kowloon
Hong Kong
Phone: 852 2766 5648
Email: stefanos.tyrovolas@polyu.edu.hk

Abstract

Background: On January 21, 2020, the World Health Organization reported the first case of severe acute respiratory syndrome coronavirus 2, which rapidly evolved to the COVID-19 pandemic. Since then, the virus has also rapidly spread among Latin American, Caribbean, and African countries.

Objective: The first aim of this study is to identify new emerging COVID-19 clusters over time and space (from January 21 to mid-May 2020) in Latin American, Caribbean, and African regions, using a prospective space–time scan measurement approach. The second aim is to assess the impact of real-time population mobility patterns between January 21 and May 18, 2020, under the implemented government interventions, measurements, and policy restrictions on COVID-19 spread among those regions and worldwide.

Methods: We created a global COVID-19 database, of 218 countries and territories, merging the World Health Organization daily case reports with other measures such as population density and country income levels for January 21 to May 18, 2020. A score of government policy interventions was created for *low*, *intermediate*, *high*, and *very high* interventions. The population’s mobility patterns at the country level were obtained from Google community mobility reports. The prospective space–time scan statistic method was applied in five time periods between January and May 2020, and a regression mixed model analysis was used.

Results: We found that COVID-19 emerging clusters within these five periods of time increased from 7 emerging clusters to 28 by mid-May 2020. We also detected various increasing and decreasing relative risk estimates of COVID-19 spread among Latin American, Caribbean, and African countries within the period of analysis. Globally, population mobility to parks and similar leisure areas during at least a minimum of implemented intermediate-level control policies (when compared to low-level control policies) was related to accelerated COVID-19 spread. Results were almost consistent when regional stratified analysis was applied. In addition, worldwide population mobility due to working during high implemented control policies and very high implemented control policies, when compared to low-level control policies, was related to positive COVID-19 spread.

Conclusions: The prospective space–time scan is an approach that low-income and middle-income countries could use to detect emerging clusters in a timely manner and implement specific control policies and interventions to slow down COVID-19 transmission. In addition, real-time population mobility obtained from crowdsourced digital data could be useful for current and future targeted public health and mitigation policies at a global and regional level.

(J Med Internet Res 2021;23(6):e22999) doi:10.2196/22999

KEYWORDS

COVID-19; transmission; digital public health; social distancing; policy; mobile data; estimate; real-time; pattern; surveillance; low and middle-income countries; emerging countries; database; surveillance

Introduction

On January 30, 2020, the World Health Organization (WHO) declared COVID-19 as a Public Health Emergency of International Concern and later characterized it as a pandemic [1]. On January 21, 2020, the WHO published the first situation report, announcing the first cases of pneumonia of unknown etiology detected in Wuhan City on December 31, 2019 [2]. By May 18, 2020, more than 200 countries reported confirmed cases of COVID-19 [2,3]. Among them, there were several Latin American, Caribbean, and African countries with limited resources to monitor, manage, and treat COVID-19. The first signs of virus spread were delayed among these regions compared with Europe, Asia, and North America [2,4].

Several government interventions have already been implemented to prevent and contain the alarming propagation of COVID-19 [5,6]. Each country has applied its own disease control measures, which vary by specific policy and implementation timing. Some countries initially implement a *lower* level of measures and policies, while others are adopting *stricter* ones. Government health and social distancing policies are evolving rapidly based on the COVID-19 transmission in each region. Policies range from traveler's temperature checks and medical screening at each country's entry point and public school closures to quarantining an entire country. Various Latin American and African countries adopted COVID-19 restriction policies rapidly [7]. Barriers such as the effectiveness of social distancing measures among low-income and middle-income countries (LMICs) have been pointed out [8]. For these reasons, global and local health policy makers and international organizations have said that the lack of health and government resources among these regions would pose barriers and challenges to halt virus spread.

Space–time surveillance is a methodology [9] that could be of use among Latin American and African regions to identify and list locations in an emergency, apply the strictest public health measures, and allocate resources. The space–time scan statistic technique is able to detect *dynamic* or emerging clusters of COVID-19, which can be used for targeted monitoring during the outbreak [10]. Since COVID-19 data are updated daily, this method could contribute to timely monitoring of the pandemic among various areas such as Latin American and African regions. In addition to statistical approaches, digital technology [11,12] could be used to understand population mobility and to assess the effectiveness of government policies or the re-evaluation of specific strategies. During the COVID-19 outbreak, smartphone software can provide information (in an anonymous way and at the country level) [13] on various characteristics of population mobility (eg, workplaces or parks). This information could be of use among countries and specifically in resource-limited settings to understand rapidly whether the government restrictions need enhancements or corrections.

In the early stage of the COVID-19 spread in the United States, prospective scan statistic methodology detected the *active* cluster in New York State, marking the area that needed specific attention [10]. To the best of our knowledge during COVID-19, except in the United States, the prospective scan method has not been applied elsewhere, though it could be a useful monitoring and intervention–decision tool. In addition, the collective effect sizes of population mobility patterns under the social distancing government policies are empirically unknown, particularly in low-income countries (LICs) and middle-income countries (MICs) with differential population vulnerability (ie, poverty, lack of resources, and health infrastructure). Given the information regarding effective treatment schemes and population vaccination going forward slowly and taking into account the delays in deliveries among countries, detection of emerging clusters among these regions will make a substantial contribution [14] to the field facilitating the translation of knowledge and implementation of evidence into COVID-19 practice and policy at the country level. In addition, it will guide authorities globally and among low- and middle-income countries (LMICs) to enhance and update if necessary the applied COVID-19 containment policies based on real-time population mobility. The first aim of this study, therefore, is to identify new emerging space–time COVID-19 clusters implementing space–time surveillance among Latin American, Caribbean, and African countries, applying a prospective space–time scan statistic methodology. This technique is a well-known method for detecting clusters of health-related events in the space–time dimension [10]. Our available data extends until May 18, 2020. Thus, we report results applying the prospective space–time scan statistic in five time periods to monitor the emerging clusters when adding updated case data: (1) January 21 to March 15, (2) January 21 to March 31, (3) January 21 to April 15, (4) January 21to April 30, and (5) January 21 to May 15. The second aim is to assess the impact of real-time population mobility patterns between January 21 and May 18 under the implemented government interventions, measurements, and policy restrictions on COVID-19 spread among Latin American, Caribbean, and African countries as well as globally. This study focuses its analysis on Latin American, Caribbean, and African countries (as a sample of LMICs) among other Asian-Pacific regions due to the excessive COVID-19 transmission in these areas. This study could serve as a learning tool presenting new information on virus surveillance and its timely detection among countries and regions with limited resources at their disposal, while population mobility patterns will facilitate public health authorities to design targeted social distancing strategies instead of *horizontal* lockdowns or social distancing measures. Results of this study in combination with lessons from other countries' experiences [15] could be helpful for policy makers at regional and international level.

Methods

Study Design

We conducted a retrospective observational longitudinal study. We obtained data on COVID-19 propagation and related risk factors from 218 countries and territories (as reported by the WHO). We compiled a data set of COVID-19 daily cases and deaths spanning the period January 21 to May 18, 2020, based on the most recent publicly available population-level information (by country), as reported by the WHO [2]. This study was approved by Parc Sanitari Sant Joan de Déu, Ethics Committee (PIC-67-20, Barcelona, Spain) and conformed to the ethical guidelines of the 1975 Declaration of Helsinki.

COVID-19 International Data

The WHO daily situation reports were used from January 21 to May 18, 2020, to assess disease transmission internationally [2]. Data on daily confirmed cases, total confirmed cases, daily confirmed deaths, total confirmed deaths, transmission classification, and time since the last reported case were compiled for 218 countries and territories. Case classifications were based on the WHO case definitions for COVID-19. Transmission was classified into three distinct groups to capture changes in the classification that the WHO applied during these 4 months: community transmission, transmission under investigation, and sporadic clusters transmission (includes sporadic transmission, clusters, and local transmission) [2]. Cases identified in cruise ships were excluded from the analysis, while cases among China's provinces were grouped together. Cases in Hong Kong, Macao, and Taiwan special administrative regions of China were classified separately since they applied different government interventions and policy measures than mainland China. Based on the WHO database, Puerto Rico was classified separately from the United States as was the case for other territories.

Countries' Government Interventions, Health Policy, and Restriction Measures

Each country's health and government policy measures were obtained as announced from each country's official source after January 21, 2020. If this was not feasible, the information was obtained from local media sources and was cross-checked with at least two sources (where possible). Additionally, two researchers cross-validated the obtained information to ensure the highest accuracy. This information was then validated using the WHO global tracking database on governments' COVID-19 response as the gold standard database [16]. Based on this information, a four-level health and government policy interventions and measures score was created, ranging from 0 to 3, which represented *low*, *intermediate*, *high*, and *very high* intervention levels [17]. These intervention and control policy categories were formed following already announced alert classification systems [18] and other international COVID-19 government response data and methodologies [19,20].

Other Baseline Assessments by Country

Index for Risk Management and World Bank Income Classification

Information regarding threat detection and risk assessment were obtained from the Index for Risk Management (INFORM) Epidemic Risk Index [21], developed by the EU Joint Research Centre in collaboration with the WHO, and was used as a measure of each country's epidemic preparedness. The INFORM index ranged from 0 to 10, and higher scores corresponded to a lower epidemic preparedness risk of a country. More detail about the development of this index can be found elsewhere [21].

The World Bank income classification system was also used to classify each of the 218 countries' income (high-income countries, upper-income countries, lower middle–income countries, and low-income countries) [22]. COVID-19 testing policy in each country was assessed as the number of days that each country started implementing COVID-19 tests in the population and as the number of days that each country implemented tracing strategies for COVID-19 cases. Information on these items was obtained from publicly available sources [20].

Cell Phones and Community Mobility Reports

The population's mobility patterns at the country level were obtained from Google community mobility reports. These reports are publicly available [23] and present movement trends over time by geography and across different place categorization such as retail and recreation places, groceries and pharmacies, parks and other similar places, transit stations, workplaces, and residential areas. Specifically, as described by Google reports, retail and recreation grouping correspond to mobility trends for places such as restaurants, cafés, shopping centers, theme parks, museums, libraries, and movie theaters. Groceries and pharmacies reports mobility trends for places such as grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies. The parks category encompasses mobility trends for places such as national parks, public beaches, marinas, dog parks, plazas, and public gardens. In addition, transit station cluster marks mobility trends for places such as public transport hubs (eg, subway, bus, and train stations). The workplace classification corresponds to the mobility trend for places of work. Finally, the residential cluster encompasses mobility trends for staying at home. These reports show how visits and length of stay at different places change compared to a specific reference period (a reference period defined by Google as, for example, January 3 to February 6, 2020). Data in these reports are generated using aggregated anonymized sets of data from users that turned on the location history setting.

Of the 218 countries and territories, 179 had complete data and were selected for the adjusted analysis. In this analysis, only COVID-19 daily new cases were analyzed. Analysis was applied globally only between real-time population mobility patterns and COVID-19 spread, while the rest of the analysis was restricted for the regions of Latin America, the Caribbean, and Africa. By May 15, 2020, we calculated the standardized incidence ratios (SIRs) [24] for 4 countries (Brazil, Peru, Uganda, and Nigeria) to compare and validate the accuracy of our prospective space–time models. COVID-19 SIR estimations were calculated as the ratio of observed counts to the expected counts.

Statistical Analysis

Based on the literature review, government interventions are not having an immediate effect on COVID-19 spread; for this reason, we considered their time scheduling based on a starting time point t0, with the addition of a seven-day lag [25]. These time lag effects only concern the modeling process via a mixed model approach, as it is when our analysis tested the government control policies.

Prospective Space–Time Scan Statistic

The early detection of emerging COVID-19 space–time clusters was determined using a prospective version of the space–time scan statistic approach [9]. The method helps to identify COVID-19 clusters in the space–time dimension, which have a significant relative risk (RR) at the end of the study period [10]. The general assumption is that the number of COVID-19 cases follows a Poisson distribution with a constant risk, which is proportional to the at-risk population of each corresponding country or territory over the geographic region under study.

This approach works using cylinders that move and scan the region of interest looking for potential space–time clusters of COVID-19 cases. The center of the cylinder is defined as the centroid of each country in the region of interest. The general working function of this technique can be summarized as follows: an unknown large number of cylinders of different spatial and temporal sizes are generated around each centroid until the maximum spatial and temporal thresholds are reached; the observed and expected case counts are computed within each cylinder, which is derived from the total number of centroids captured in each cylinder.

In this manner, the RR is defined as having more observed than expected COVID-19 cases within each cylinder. We determined the elevated RR of COVID-19 calculating maximum log-likelihood ratio tests. Furthermore, 100 runs of Monte Carlo testing were used to depict the empirical distribution of the log-likelihood ratio, assuming constant risk. This distribution allows us to assess the statistical significance of space–time clusters (*P* value <.05), and the cylinder with the largest log-likelihood ratio is the most likely cluster.

In our study, we reported the significant emerging clusters of COVID-19 at the country level for the Latin American, Caribbean, and African regions, and computed the estimated RR, which identifies the risk for the population to COVID-19 within a cluster in comparison with the risk outside of the cluster. Moreover, five time periods were monitored: (1) January 21 to March 15; (2) January 21 to March 31; (3) January 21 to April 15; (4) January 21 to April 30; and (5) January 21 to May 15, 2020.

This analysis was carried out using the R package *scanstatistics* in R version 3.6.3 (R Foundation for Statistical Computing) and follows previous prospective scan statistic work [9,10,26].

Mixed Models Analysis

We also fitted a negative binomial (NB) mixed model, with daily new COVID-19 cases as the outcome. The model accounts for a linear trend with respect to time since the appearance of the first COVID-19 case, taking into account the varying secular trends across regions and the treatment–effect heterogeneity across time, and was adjusted for the countries' preparedness in epidemics (INFORM index), COVID-19 testing policy, COVID-19 type of transmission for each country through time, populations' real-time mobility patterns, their interaction with each level of government control policy, and the country's income level. In this mixed model, all the predictors are assumed to be fixed effects; however, the intercept includes a country-level random effect term. As offset, the natural logarithm of the total population was added to the generalized linear predictor function of the NB component to account for the variable number of daily new COVID-19 cases per country population. Models were tested globally and regionally (for Latin American, Caribbean, and African countries). The maximum likelihood estimation procedure was used to fit all multilevel analysis models. In this mixed model, only three out of five real-time mobility patterns for workplaces, parks, and similar places as well as mobility for food and drug supplies were applied to avoid collinearity, as the correlation among the mobility variables was higher than 0.70. Mixed model analysis was carried out using the R package g*lmmTMB* in R Version 3.6.3 [27].

Validation Analysis

First, the predicted validity of the prospective scan statistic model was tested evaluating the relation between the RRs of the prospective scan statistic and the SIR estimations using Pearson rho coefficients. Second, we used the Man-Whitney *U* test to explore the increase in mean COVID-19 cases before and after April 15, 2020, and then, we checked if the prospective scan had predicted a potential COVID-19 cluster. We followed classical reported criteria to classify a correlation as weak (≤0.3), moderate (0.4-0.6), and strong (≥0.7; coefficients are presented as absolute values) [28]. All *P* values are based on two-sided tests. A *P* value ≤.05 was considered as significant.

Results

Prospective Scan Statistics and Emerging Country-Level Results Between January 21 and May 15 for Latin America, the Caribbean, and Africa

Tables 1 and 2 provide the characteristics of the significant COVID-19 emerging space–time clusters at the country level among LMICs, from January 21 to May 15, 2020. Analyzing COVID-19 spread between January 21 to March 15, 7 major clusters among Latin American, Caribbean, and African countries were revealed. For the Latin American and Caribbean countries, cluster 1 included 15 countries with a RR>1 (ie, having more observed than expected COVID-19 cases). Saint Kitts and Nevis (cluster 3) showed an extremely elevated RR of 19.31 (*P*<.001). Exploring the African region during the same period, cluster 4 integrated most of the countries with a RR of 9.46 (*P*<.001), and cluster 5 encompassed Madagascar, Mauritius, Mayotte, Reunion, and Seychelles with a RR of 21.35 (*P*<.001). Rwanda and Uganda were grouped in cluster 6 and Sudan in cluster 7. Both these clusters marked the most elevated RR (cluster 6: RR 37.75, *P*<.001; cluster 7, RR 45.75, *P*<.001; Figures S1 and S2 in Multimedia Appendix 1).

Using data for the period of January 21 to March 31, 2020, allowed us to assess the evolution of COVID-19 spread among LMICs. It was shown that the initial 7 emerging COVID-19 country clusters among Latin America, the Caribbean, and Africa, when the period of analysis extended for 15 days, were spread to 20 clusters. Cluster 5 that included only Antigua and Barbuda Island had the most elevated relative risk (RR 60.48; *P*<.001) followed by cluster 6 (Puerto Rico, Saint Maarten, and Virgin Islands) with a RR of 23.27 (*P*<.001) and cluster 8 (Dominica) with a RR of 21.85 (*P*<.001; Figures S3 and S4 in Multimedia Appendix 1).

Analyzing data from January 21 to April 15-30 and to May 15, 2020, it was shown that the spread was further extended with the evolution of time. Specifically, when the period of analysis extended to April 15, the virus spread was reported to 27 clusters among Latin America, the Caribbean, and Africa, and by April 30 and May 15, the emerging clusters were 29 and 28, respectively. For the period between January 21 to April 15 for Latin America and the Caribbean, the cluster with the highest RR was Chile (RR 18.02; *P*<.001), followed by Antigua and Barbuda (RR 14.22; *P*<.001), and Mexico (RR 14.07; *P*<.001). For Africa, the emerging clusters were Djibouti (RR 165.84; *P*<.001), followed by Mauritius (RR 136.05; *P*<.001) and Egypt (RR 52.20; *P*<.001). Focusing on the period between January 21 to April 30 and to May 15, we showed extended COVID-19 spread following similar patterns as previously mentioned (Figures S5-S10 in Multimedia Appendix 1).

Table 1. Emerging COVID-19 space–time clusters and their RR for having more observed than expected COVID-19 cases, from January 21 to March 15 and March 31, 2020, at the country level within the Latin American, Caribbean, and African regions.

Region, date range, and cluster number	Cluster	Duration	RR[a,b]
Latin America and Caribbean			
January 21 to March 15			
1	Argentina, Bolivia, Brazil, Barbados, Colombia, Grenada, Guyana, Saint Lucia, Peru, Paraguay, Suriname, Trinidad, Uruguay, Saint Vincent, Venezuela	March 11-15, 2020	3.02
2	Bahamas, Belize, Costa Rica, Cuba, Cayman Islands, Ecuador, Guatemala, Honduras, Haiti, Jamaica, Mexico, Nicaragua, Panama, El Salvador, Turks and Caicos Islands	March 9-15, 2020	4.04
3	Saint Kitts and Nevis	March 12-15, 2020	19.31
January 21 to March 31			
1	Bolivia, Colombia, Ecuador, Peru	March 17-31, 2020	11.61
2	Bahamas, Belize, Costa Rica, Cuba, Cayman Islands, Dominican Republic, Guatemala, Honduras, Haiti, Jamaica, Mexico, Nicaragua, Panama, El Salvador, Turks and Caicos Islands	March 22-31, 2020	3.87
3	Barbados, Grenada, Trinidad, Saint Vincent	March 22-31, 2020	2.53
4	Argentina, Chile, Paraguay, Uruguay	March 24-31, 2020	2.79
5	Antigua and Barbuda	March 21-31, 2020	60.48
6	Puerto Rico, Saint Maarten, Virgin Islands	March 20-31, 2020	23.27
7	Curacao	March 27-31, 2020	2.99
8	Dominica	March 22-31, 2020	21.85
9	Saint Lucia	March 26-31, 2020	1.25
Africa			
January 21 to March 15			
4	Benin, Burkina Faso, Ivory Coast, Algeria, Ghana, Guinea, Gambia, Mali, Mauritania, Senegal, Sierra Leone, Togo	March 6-15, 2020	9.46
5	Madagascar, Mauritius, Mayotte, Reunion, Seychelles	March 13-15, 2020	21.35
6	Rwanda, Uganda	March 9-15, 2020	37.75
7	Sudan	March 13-15, 2020	45.75
January 21 to March 31			
10	Madagascar, Mozambique, Mauritius, Mayotte, Reunion, Seychelles	March 17-31, 2020	8.75
11	Rwanda, Uganda	March 18-31, 2020	34.34
12	Benin, Burkina Faso, Ivory Coast, Algeria, Ghana, Guinea, Gambia, Mali, Mauritania, Senegal, Sierra Leone, Togo	March 19-31, 2020	3.83
13	Botswana	March 15-31, 2020	6.92
14	Equatorial Guinea	March 21-31, 2020	5.91
15	Djibouti	March 22-31, 2020	31.00
16	Egypt, Sudan	March 20-31, 2020	4.47
17	Guinea-Bissau	March 30-31, 2021	3.27
18	Zambia	March 26-31, 2020	13.84
19	Tanzania	March 18-31, 2020	2.85
20	Morocco	March 28-31, 2021	2.86

[a] RR: relative risk estimate.

[b] All RRs have a *P* value <.001

Table 2. Emerging COVID-19 space–time clusters and their RR for having more observed than expected COVID-19 cases from January 21 to April 15, April 30 and May 15, 2020, at the country level within the Latin American, Caribbean, and African regions.

Region, date range, and cluster number	Cluster	Duration	RR[a,b]
Latin America and Caribbean			
January 21 to April 15			
1	Colombia, Costa Rica, Ecuador, Panama, Peru	March 29-April 15, 2020	5.27
2	Barbados	March 31-April 15, 2020	2.81
3	Mexico	March 21-April 15, 2020	14.07
4	Dominican Republic	March 22-April 15, 2020	5.94
5	Puerto Rico, Saint Maarten, Virgin Islands	March 22-April 15, 2020	9.88
6	Antigua and Barbuda	March 21-April 15, 2020	14.22
7	Saint Vincent	March 17-April 15, 2020	2.08
8	Chile	March 29-April 15, 2020	18.02
9	Curacao	April 6-15, 2020	1.94
10	Belize, Guatemala, El Salvador	March 22-April 15, 2020	4.54
11	Bahamas, Cuba	March 18-April 15, 2020	1.74
12	Saint Lucia	April 10-15, 2020	1.33
13	Argentina, Uruguay	April 12-15, 2020	1.61
14	Saint Kitts and Nevis	April 14-15, 2020	4.66
15	Dominica	March 22-April 15, 2020	3.98
January 21 to April 30			
1	Colombia, Ecuador, Panama, Venezuela	April 7-30, 2020	5.85
2	Barbados	April 15-30, 2020	3.44
3	Mexico	March 21-April 30, 2020	7.64
4	Dominican Republic	March 29-April 30, 2020	3.84
5	Antigua and Barbuda, Dominica, Saint Kitts and Nevis, Saint Lucia	April 22-30, 2020	1.67
6	Peru	March 29-April 30, 2020	2.95
7	Saint Maarten	March 29-April 30, 2020	11.47
8	Chile	March 29-April 30, 2020	6.47
9	Virgin Islands	March 20-April 30, 2020	2.94
10	El Salvador	March 17-April 30, 2020	4.49
11	Costa Rica	April 28-30, 2020	1.26
January 21 to May 15			
1	Bolivia, Brazil, Barbados, Colombia, Grenada, Guyana, Peru, Paraguay, Suriname, Trinidad, Uruguay, Venezuela	April 25-May 15, 2020	3.40
2	Panama	April 14-May 15, 2020	8.28
3	Ecuador	April 7-May 15, 2020	5.32
4	Mexico	March 24-May 15, 2020	4.54
5	Dominican Republic	April 8-May 15, 2020	2.80
6	Saint Lucia	May 6-15, 2020	1.64
7	Costa Rica	May 8-15, 2020	1.34
8	Saint Maarten	March 29-May 15, 2020	4.41
9	Jamaica	May 4-15, 2020	1.35
10	Chile	March 29-May 15, 2020	3.17

Region, date range, and cluster number	Cluster	Duration	RR[a,b]
11	Antigua and Barbuda	March 21-May 15, 2020	1.85
Africa			
January 21-April 15			
16	Botswana, Mozambique, Malawi, Swaziland, Zambia, Zimbabwe	March 22-April 15, 2020	5.91
17	Benin, Burkina Faso, Ivory Coast, Algeria, Ghana, Guinea, Gambia, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Senegal, Sierra Leone, Togo	March 31-April 15, 2020	5.54
18	Rwanda, Uganda	March 19-April 15, 2020	15.17
19	Djibouti	April 4-15, 2020	165.84
20	Mauritius	March 20-April 15, 2020	136.05
21	Gabon, Equatorial Guinea	March 23-April 15, 2020	5.17
22	Reunion	April 14-15, 2020	37.35
23	Egypt	April 13-15, 2020	52.20
24	Sao Tome and Principe	April 12-15, 2020	5.56
25	Libya, Tunisia	April 14-15, 2020	3.47
26	Somalia	March 14-15, 2020	16.18
27	Sudan	March 22-April 15, 2020	1.84
January 21 to April 30			
12	Benin, Burkina Faso, Ivory Coast, Algeria, Ghana, Guinea, Gambia, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Senegal, Sierra Leone, Togo	March 31-April 30, 2020	4.26
13	Comoros, Mozambique, Malawi, Swaziland, Zambia, Zimbabwe	March 22-April 30, 2020	4.85
14	Djibouti	April 5-30, 2020	108.36
15	Mauritius	March 20-April 30, 2020	117.76
16	Cameroon, Gabon, Equatorial Guinea, Sao Tome and Principe	April 5-30, 2020	4.71
17	Rwanda, Uganda	March 19-April 30, 2020	6.57
18	Egypt	April 13-30, 2020	33.32
19	Nigeria	April 28-30, 2020	3.83
20	Tunisia	April 28-30, 2020	4.03
21	Reunion	April 11-30, 2020	4.05
22	Botswana	March 20-April 30, 2020	1.88
23	Chad	April 24-30, 2020	2.69
24	Democratic Republic of the Congo	April 26-30, 2020	6.08
25	South Sudan	April 16-30, 2020	2.57
26	Sudan	April 23-30, 2020	2.61
27	Cape Verde	April 29-30, 2020	2.16
28	Republic of Kong	April 29-30, 2020	3.42
29	Somalia	March 14-April 30, 2020	6.10
January 21-May 15			
12	Benin, Burkina Faso, Ivory Coast, Algeria, Ghana, Guinea, Gambia, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Senegal, Sierra Leone, Togo	April 3-May 15, 2020	4.84
13	Comoros, Mozambique, Malawi, Swaziland, Zambia, Zimbabwe	April 11-May 15, 2020	7.35
14	Mauritius	March 20-May 15, 2020	151.42

Region, date range, and cluster number	Cluster	Duration	RR[a,b]
15	Djibouti	April 4-May 15, 2020	60.47
16	Cameroon, Gabon, Equatorial Guinea, Sao Tome and Principe	April 12-May 15, 2020	5.65
17	Egypt	April 13-May 15, 2020	25.85
18	Nigeria	April 28-May 15, 2020	2.33
19	Chad	April 28-May 15, 2020	4.38
20	Rwanda, Uganda	March 19-May 15, 2020	3.79
21	Democratic Republic of the Congo	May 8-15, 2020	11.98
22	South Sudan, Tanzania	May 3-15, 2020	4.35
23	Reunion	April 14-May 15, 2020	3.68
24	Republic of Congo	May 4-15, 2020	3.45
25	Seychelles	May 6-15, 2020	3.31
26	Cape Verde	May 5-15, 2020	1.49
27	Libya	April 7-May 15, 2020	2.12
28	Botswana	March 20-May 15, 2020	1.26

[a] RR: relative risk estimate.

[b] All RRs have a *P* value <.001.

Predictive Validity of the Prospective Scan Model

To evaluate the predictive validity of the prospective scan statistic approach, we used Brazil, Peru, Uganda, and Nigeria as country examples. Figure S11 in Multimedia Appendix 1 illustrates the RRs of the prospective scan approach and the SIR estimations. The correlation of both models for all 4 countries was strong and significant (Uganda: ρ=0.78, *P*=.04; Nigeria: ρ=0.98, *P*=.002; Brazil: ρ=0.95, *P*=.01; and Peru: ρ=0.86, *P*=.03). In addition, among all 4 countries that prospective scan modelling predicted a potential outbreak after April 15, 2020, we found a significant increase in the comparison of the COVID-19 mean new cases before and after April 15 (Uganda: W=697, *P*=.008; Nigeria: W=1280, *P*<.001; Brazil: W=1596, *P*<.001; and Peru: W=1246, *P*<.001).

COVID-19 Spread in Relation to Real-time Population Mobility Patterns Between January 21 and May 18 Globally and Regionally for Latin America, the Caribbean, and Africa

Population Mobility Patterns and COVID-19 Spread at the Global Level

COVID-19 daily new cases and real-time population mobility changes by region are presented in Figure S12 in Multimedia Appendix 1. Among the 3 population mobility patterns, a reduced change in comparison with the reference period of time was observed. Population mobility patterns and COVID-19 spread worldwide and by the Latin American, Caribbean, and African region are reported in Table 3. Worldwide, between January and May 2020, population mobility to all kinds of food places, drug stores, and pharmacies was not associated with COVID-19 spread, while the use of park places (ie, national and city parks, public beaches, and dog parks) and mobility to workplaces were negatively related with COVID-19 spread (mobility to parks and similar places: *b*=–0.03, 95% CI –0.04 to –0.02); workplaces mobility: *b*=–0.03, 95% CI –0.05 to –0.02). However, when the interaction effect between government control policies (intermediate, high, and very high) and population's mobility patterns was applied and compared with that of low-level government control policies, different trends were extracted regarding COVID-19 spread. It was observed that COVID-19 spread changes significantly throughout mobility confounders depending on the degree of the implemented control policies. Specifically, with the implementation of intermediate, high, and very high control policies, mobility to parks and other similar places (like beaches, dog parks, and others) were related with increased COVID-19 spread (*b*=0.02, 95% CI 0.01-0.03; high-level interventions: *b*=0.02, 95% CI 0.01-0.03; very high–level interventions: *b*=0.02, 95% CI 0.01-0.03) when compared with the population mobility in parks during the implementation of low-level government control policies. Similar increased COVID-19 spread estimates were shown for population mobility to workplaces during high-level and very high–level movement restrictions (high-level interventions: *b*=0.02, 95% CI 0.01-0.04; very high–level interventions: *b*=0.03, 95% CI 0.01-0.04) when compared with population mobility to workplaces during the implementation of low-level government policies.

Table 3. Results from mixed model analysis that evaluated the COVID-19 spread with government interventions, their interaction with population mobility patterns, and other factors during the first 4 months of the outbreak.[a]

Items	Global, *b* (95% CI)	Latin America, the Caribbean, and Africa, *b* (95% CI)	Latin America and the Caribbean, *b* (95% CI)	Africa, *b* (95% CI)
Number of days since first case	–0.003 (–0.01 to 0.01)	0.03 (0.00[b] to 0.06)	0.09 (0.05 to 0.13)	0.02 (–0.05 to 0.09)
Low-level interventions (reference category)	N/A[c]	N/A	N/A	N/A
Intermediate-level interventions	–0.90 (–1.23 to –0.56)	0.48 (–0.62 to 1.58)	2.42 (–1.50 to 6.35)	0.41 (–1.20 to 2.03)
High-level interventions	0.88 (0.59 to 1.16)	0.92 (0.35 to 1.49)	2.42 (1.52 to 3.32)	0.90 (–0.19 to 1.98)
Very high–level interventions	0.74 (0.40 to 1.08)	0.27 (–0.39 to 0.94)	1.94 (0.73 to 3.16)	0.25 (–0.88 to 1.39)
Population mobility for food and drug supplies	0.004 (–0.02 to 0.02)	–0.01 (–0.05 to 0.03)	–0.004 (–0.05 to 0.05)	–0.03 (–0.10 to 0.05)
Population mobility to parks/leisure activities	–0.03 (–0.04 to –0.02)	–0.07 (–0.10 to –0.05)	–0.10 (–0.13 to –0.07)	–0.04 (–0.10 to 0.02)
Population mobility to workplace	–0.03 (–0.05 to –0.02)	–0.02 (–0.04 to 0.00[b])	–0.003 (–0.02 to 0.01)	–0.05 (–0.09 to –0.01)
Number of days since first case × low-level interventions (reference category)	N/A	N/A	N/A	N/A
Number of days since first case × intermediate-level interventions	0.04 (0.03 to 0.05)	–0.02 (–0.07 to 0.03)	–0.28 (–0.92 to 0.36)	0.0004 (–0.09 to 0.09)
Number of days since first case × high-level interventions	–0.001 (–0.01 to 0.01)	0.002 (–0.03 to 0.03)	–0.06 (–0.10 to –0.02)	0.01 (–0.06 to 0.08)
Number of days since first case × very high–level interventions	0.005 (–0.01 to 0.01)	–0.001 (–0.03 to 0.03)	–0.07 (–0.11 to –0.02)	0.006 (–0.06 to 0.07)
Low-level intervention × population mobility for food and drug supplies (reference category)	N/A	N/A	N/A	N/A
Intermediate-level intervention × population mobility for food and drug supplies	0.02 (–0.00 to 0.04)	0.02 (–0.05 to 0.09)	0.01 (–0.08 to 0.11)	0.05 (–0.09 to 0.18)
High-level intervention × population mobility for food and drug supplies	0.004 (–0.02 to 0.02)	0.03 (–0.00[b] to 0.07)	0.03 (–0.03 to 0.08)	0.03 (–0.05 to 0.11)
Very high–level intervention × population mobility for food and drug supplies	0.001 (–0.02 to 0.02)	0.02 (–0.02 to 0.05)	–0.003 (–0.05 to 0.05)	0.05 (–0.03 to 0.12)
Low-level intervention × population mobility to visit parks and do leisure activities (reference category)	N/A	N/A	N/A	N/A
Intermediate-level intervention × population mobility to visit parks and do leisure activities	0.02 (0.01 to 0.03)	0.03 (–0.01 to 0.06)	–0.02 (–0.28 to 0.23)	–0.02 (–0.10 to 0.06)
High-level intervention × population mobility to visit parks and do leisure activities	0.02 (0.01 to 0.03)	0.05 (0.02 to 0.07)	0.07 (0.04 to 0.10)	0.04 (–0.02 to 0.11)
Very high–level intervention × population mobility to visit parks and do leisure activities	0.02 (0.01 to 0.03)	0.05 (0.03 to 0.07)	0.09 (0.06 to 0.13)	–0.001 (–0.06 to 0.06)
Low-level intervention × population mobility to workplaces (reference category)	N/A	N/A	N/A	N/A
Intermediate-level intervention × population mobility to workplaces	0.01 (–0.01 to 0.03)	0.03 (–0.03 to 0.08)	0.03 (–0.10 to 0.16)	0.07 (–0.02 to 0.16)
High-level intervention × population mobility to workplaces	0.02 (0.01 to 0.04)	0.01 (–0.01 to 0.03)	0.0004 (–0.02 to 0.02)	0.04 (–0.00[b] to 0.08)
Very high–level intervention × population mobility to workplaces	0.03 (0.01 to 0.04)	0.02 (–0.00[b] to 0.03)	–0.0002 (–0.02 to 0.02)	0.04 (–0.00[b] to 0.08)

[a] Models were also adjusted by country income level; preparedness in epidemics (Index for Risk Management); COVID-19 type of transmission (ie, community transmission or local transmission); and COVID-19 testing and tracing policies (in days).

[b] These are values less than 0.005 in absolute numbers.

[c] N/A: not applicable.

Population Mobility Patterns and COVID-19 Spread for Latin America, the Caribbean, and Africa

When the analysis was stratified by Latin American, Caribbean, and African countries, specific trends in COVID-19 spread were shown due to distinct population's mobility patterns (Table 3). As noted, the coefficients of the interaction between the populations' mobility to parks and similar places and social distancing measures were consistent at the regional level, apart from the region of Africa. Specifically, when high-level and very high–level control policies were applied in comparison with those at a low level, only people's mobility to parks and similar places was related with increased COVID-19 spread (exception was the African countries where the results were not significant; ie, Latin America, the Caribbean, and Africa: high-level interventions concurrently with mobility to parks and similar places *b*=0.05, 95% CI 0.02-0.07; very high–level interventions concurrently with mobility to parks and similar places *b*=0.05, 95% CI 0.03-0.07). Moreover, the interaction between populations' mobility to workplaces and social distancing measures showed heterogeneity among the tested regions. However, apart from the global analysis, results were not significant for the Latin American, Caribbean, or African regions.

Finally, we conducted a sensitivity analysis to assess our inferences for large countries in terms of area extension (due to possible subnational mitigation policies) and to avoid a possible bias effect toward big country areas, taking into account our analytical sample of countries (n=179). For the sensitivity analysis, we fitted all models again for data on COVID-19 spread removing the 5 top countries with the largest area worldwide (ie, Russia, Canada, China, the United States, and Brazil) and the top 3 countries with the largest area regionally for Latin America and for Africa (Brazil, Angola, Argentina, etc). Again, the observed results remained in the same direction at the global and regional level, as was previously mentioned (data not shown in the text).

Discussion

Principal Findings

This study analyzed the geographical and temporal COVID-19 spread among LMICs in Latin America, the Caribbean, and Africa using the prospective space–time scan statistical methodology and the impact of real-time population mobility patterns during the implemented government interventions in the area of interest between January 21 and May 18, 2020. First, analyzing the current data with scan statistics at five prospective time periods, it was shown that virus spread was rapid and at alarming rates since March 15, where we detected 7 emerging COVID-19 clusters, which at May 15 had spread to 28, among the regions of Latin America, the Caribbean, and Africa. As governments decide their strategies in response to the pandemic, surveillance is of importance especially among LMICs that have limited resources at their disposal; hence, the prospective scan statistic could be used as a useful surveillance tool at the international, national, and subnational levels. Second, as presented in 4 country-specific examples, the prospective scan statistic showed high predictive validity with classic surveillance technics. Third, when the real-time mobility to parks, beaches, and other similar places as well as the mobility to workplaces were tested as individual factors, it was shown that these patterns were related with reduced COVID-19 spread. However, worldwide, the population movement to parks, beaches, and other similar places (although more reduced than the reference period) seemed to be related with increased virus spread with all levels (intermediate, high, very high) of government control policies activated (when compared with the low-level government control policies). Fourth, similar trends were shown for population mobility to workplaces when high and very high–level control policies (after comparing them with low-level government interventions) were implemented worldwide. Fifth, stratified analysis for the Latin American, Caribbean, and African regions showed a variety of patterns mostly following the entire samples' tendency (ie, real-time mobility to parks when the social distancing measures were implemented). Governments are applying social and mobility restriction measures to slowdown the COVID-19 spread, but there is limited information about the real-time population mobility patterns, and based on our analysis, this information could help public health authorities to design effective strategies to slow down virus transmission.

The major strength of the prospective space–time scan methodology is the ability to add dynamically updated data sets and reapply the analysis to extract new emerging COVID-19 clusters, while it also has the ability to monitor the growing or shrinking COVID-19 evolution among initial detected clusters. Our analysis showed that Antigua and Barbuda had an emerging COVID-19 cluster with one of the highest RRs in Latin America and the Caribbean by the end of March 2020 and continues with a shrinking magnitude by mid-May. Similar trends were reported for Mexico and Chile. The same tool could be used subnationally among these countries to detect emerging clusters at the cross-national level (as has been done for the United States [10]). Regarding African regions, Djibouti, Mauritius, and Egypt showed growing and reducing magnitudes in COVID-19 spread from January 21 to May 15, 2020. This kind of information could be helpful to the relative stakeholders since it gives the opportunity to the public health authorities to evaluate constantly the effectiveness of the implemented mitigation and control strategies. Our comparative analysis between prospective scan and SIR modeling among 4 areas showed similar predictive results in virus spread. As has recently been reported, effective COVID-19 surveillance and monitoring need to include additional information on suspected, probable, and negative COVID-19 tests for a holistic understanding of COVID-19 transmission patterns [29], something that is often not possible for LMICs and could be marked as a barrier. Thus, future studies, data sets, and research funding are needed [30]. Health policy research showed that countries should not phase out social distancing policies until they establish strong systems that could effectively monitor the COVID-19 spread [15]. For this reason, at the early stages of the virus spread, the prospective scan methodology could serve as a useful public health surveillance tool especially among LMICs that are facing substantial limitations to monitor and detect virus transmission.

A lot of discussion has happened about the role of government interventions and control policies on COVID-19 spread. Until today, there is limited information on the impact of social restrictions and control policies at the global, regional, and national levels [15,25]. In Europe, the strictest government policy measures are related with less virus spread [31]. Targeted national health policies with effective screening and isolation tools as well as support systems are needed [15]. Based on that, real-time population mobility patterns during this intervention and social distancing period could be useful to stakeholders and policy makers to plan current and future public health–targeted strategies. Recent studies showed the important role of COVID-19–targeted strategies at the national level [32,33]. Moreover, this kind of information could serve among LMICs with limited resources in social distancing implementation, allowing them to plan targeted mobility control activities [34].

To date, although this kind of information is publicly available from well-known crowdsource applications, analyses at a national, regional, or global level are lacking. Our study used the publicly available crowdsource mobility information and showed that at the global level, when all kinds of government control policies were implemented, mobility to parks, beaches, and other leisure places was related with accelerated COVID-19 spread when compared with places where low-level control policies were effective. At a global level, peoples' mobility for work was also related to increased virus spread when high-level and very high–level government interventions were active, after comparing with the reference category of low-level government interventions. However, at the regional level (LMICs among Latin America, the Caribbean, and Africa), the results were not significant, showing that the aforementioned relation is mainly driven from the rest of the areas around the world. At this point, it has to be noted that, when we did not take into account the interaction with governmental social distancing interventions, the aforementioned mobility patterns were related with reduced virus transmission.

During the period that a region is facing increased dynamics of COVID-19 spread, social activities and engagement are associated with increased risk for virus spread [35]. For this reason, the WHO and other public health organizations are recommending avoiding crowded conditions [36]. Recent data from the United States showed that lower mobility to workplace and retail locations is related with lower virus transmission [35]. Our entire sample analysis showed a positive relation between mobility to places like parks and workplaces while governments applied social distancing measures. Similar findings were also reported by other researchers for the United States [35]. In addition, our regional stratified analysis showed consistent findings with the global one, except from the region of Africa. Taking into account that human mobility is a complicated concept (and at this point is analyzed collectively by using crowdsource data), we may hypothesize that individuals' behavior (use of face mask) [32], people's dynamic network [33] when visiting these places, and seasonality could be potential explaining mechanisms of virus spread [35,37]. The aforementioned findings could guide stakeholders on specific social distancing implementations and enforcement planning [38]. In the past months, countries are introducing various nonpharmaceutical intervention strategies in their local health policy agendas. These results can be used as *roadmap* indicators for specific social distancing planning. Targeted implemented policies could lead to further suppressed levels of virus spread, with less negative effects on the economy and citizens [39]. To this extent, a recent study noted that, during the phasing out of government social distancing policies, higher mobility at workplaces was correlated with increasing virus spread [40].

Effectiveness of the Current Measures and Current and Future Challenges

To date, countries have adopted divergent restriction strategies to suppress and halt COVID-19 transmission. Stricter social distancing policies seem effective in suppressing virus spread [31]. Differences in innovative surveillance techniques, virus transmission monitoring, COVID-19 cases tracing, systematic population testing, and isolating practices have been shown [15]. Regions with previous experiences in infectious diseases (eg, severe acute respiratory syndrome) have invested in their public health care system's reformation to efficiently handle the current outbreak [15]. Western societies (eg, the European Union and the United States) seem to lack this kind of planning [41]. In addition, recent studies showed that COVID-19 spread could be more rapid among more prosperous countries [15,42]. Countries need to organize their health systems [15,41], establishing effective infectious diseases and crisis management planning [43] (eg, enhance monitoring techniques and screening tools) to prevent virus spread in the community. Future longitudinal studies may be needed to better describe the relation of real-time mobility data with COVID-19 transmission.

Strengths and Limitations

This is among the first studies using COVID-19 prospective surveillance analysis among LMICs, exploring COIVD-19 spread in relation with real-time population mobility patterns. However, this study shares common limitations with previous studies of this kind [17,25]. Specifically, there were challenges in capturing uncertainty (completeness of the WHO COVID-19 data set or government interventions being announced on one day but only being applied after several days) and lags in data availability, which may not fully capture temporal trends of COVID-19 spread. We extracted only mobility patterns from smartphones using Google software to ensure homogeneity of the used information. In addition, this study had only the ability to analyze data from regions in which mobile phone information was obtainable [35]. For example, the use of crowdsourcing digital data for the extraction of real-time population trends through mobile phones may be limited particularly in Africa (data for selected African countries reported that about one-third of adults own smartphones) [44]. This may have altered the findings of this study. In addition, certain large countries applied subnational control policies at different time points, which could have affected our findings. In that manner, we applied a sensitivity analysis excluding large countries from the global and regional sample, and testing whether those countries had an impact on the inferential analyses. The applied aforementioned analysis showed that the results remained similar. Additionally, our investigation focused on data variations in the COVID-19 spread from January to May 2020.

Therefore, the results of this paper should be interpreted with caution, as they only relate to the underlying data collection conditions and period. As COVID-19 is an infection with dynamic transmission and all the variables we use may variate in the future, we do not think it would be appropriate to make conclusions beyond May, as further data and analysis would be required. To this extent, it should also be noted that this mixed model analysis assumes that the impact of each relative mobility pattern change has the equal relative impact among countries and across time (as an additional adjustment to this extent was not possible). Next, although our study adjusted for various confounders, we could not consider physical distancing recommendations (ie, 1 or 2 meters) or other precautionary measures and conditions due to lack of data. In addition, some of the mobility variables used in this analysis (eg, parks) may be also affected from weather seasonality. The prospective space–time scan statistic used case data for confirmed cases, so suspected and probable cases were not considered due to the unavailability of the WHO COVID-19 data set. In addition, the prospective scan methodology does not allow for adjustment of age and other covariates. These limitations may alter the true magnitude of the COVID-19 spread as presented by using the prospective scan statistic.

Conclusions

We used publicly available WHO daily reports to identify emerging space–time clusters of COVID-19 at the country level among Latin America, the Caribbean, and Africa for five separate time periods. It was shown that the prospective scan is a tool that LMICs could use to detect emerging clusters and implement specific control policies and interventions to slowdown COVID-19 transmission. In addition, we found that different kinds of real-time population mobility patterns were related with different magnitudes of COVID-19 spread worldwide. The findings of this study give insights that may help in COVID-19 screening and detection strategies as well as in government–specific COVID-19 control planning.

Acknowledgments

All data generated or analyzed during this study are included in this paper and its supplementary information files.

ST was supported by the Foundation for Education and European Culture, the Miguel Servet programme (reference CP18/00006), and the Fondos Europeos de Desarrollo Regional. DF is a Serra Húnter Fellow and was supported by Marsden grant E2987-3648 administrated by the Royal Society of New Zealand, and by grant 2017 SGR 622 (GRBIO) administrated by the Departament d'Economia i Coneixement de la Generalitat de Catalunya (Spain). WP was supported by NASA-ROSES Grant NNX15AP74G.

We would like to thank Dr Somanth Chatterji for providing useful suggestions on this study.

Authors' Contributions

ST, MM, IGV, DF, and WP designed the study and drafted the manuscript. ST, MM, and IGV obtained the data from publicly available sources. ST, IGV, and DF analyzed the data. All authors contributed to the interpretation of the data and revision of the manuscript. All authors had primary responsibility for final content and act as guarantors. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 11368 KB - jmir_v23i6e22999_app1.docx]

References

1. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). World Health Organization. 2020 Jan 30. URL: https://tinyurl.com/2c6x9psn [accessed 2020-06-24]
2. Novel coronavirus (2019-nCoV) situation reports. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-04-11]
3. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
4. Davis S. Coronavirus: newspaper round-up after COVID-19 epicentre shifts to Europe. Euronews. URL: https://www.euronews.com/2020/03/20/coronavirus-newspaper-round-up-after-covid-19-epicentre-shifts-to-europe [accessed 2020-04-11]
5. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA 2020 May 19;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]

6. Gupta R, Dhamija R. Covid-19: social distancing or social isolation? BMJ 2020 Jun 18;369:m2399. [doi: 10.1136/bmj.m2399] [Medline: 32554394]
7. Burki T. COVID-19 in Latin America. Lancet Infect Dis 2020 May;20(5):547-548 [FREE Full text] [doi: 10.1016/S1473-3099(20)30303-0] [Medline: 32311323]
8. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boëlle PY, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. Lancet 2020 Mar 14;395(10227):871-877 [FREE Full text] [doi: 10.1016/S0140-6736(20)30411-6] [Medline: 32087820]
9. Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Med 2005 Mar;2(3):e59 [FREE Full text] [doi: 10.1371/journal.pmed.0020059] [Medline: 15719066]
10. Desjardins M, Hohl A, Delmelle E. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. Appl Geogr 2020 May;118:102202 [FREE Full text] [doi: 10.1016/j.apgeog.2020.102202] [Medline: 32287518]
11. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. Lancet Digit Health 2020 Apr;2(4):e201-e208 [FREE Full text] [doi: 10.1016/S2589-7500(20)30026-1] [Medline: 32309796]
12. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in mainland China: observational infoveillance study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
13. Ye J. The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. JMIR Med Inform 2020 Jul 14;8(7):e19866 [FREE Full text] [doi: 10.2196/19866] [Medline: 32568725]
14. Tariq A, Lee Y, Roosa K, Blumberg S, Yan P, Ma S, et al. Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020. BMC Med 2020 Jun 03;18(1):166 [FREE Full text] [doi: 10.1186/s12916-020-01615-9] [Medline: 32493466]
15. Han E, Tan MMJ, Turk E, Sridhar D, Leung GM, Shibuya K, et al. Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. Lancet 2020 Nov 07;396(10261):1525-1534 [FREE Full text] [doi: 10.1016/S0140-6736(20)32007-9] [Medline: 32979936]
16. Tracking public health and social measures: a global dataset. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/phsm [accessed 2021-02-16]
17. Fernández D, Giné-Vázquez I, Liu I, Yucel R, Nai Ruscone M, Morena M, et al. Are environmental pollution and biodiversity levels associated to the spread and mortality of COVID-19? A four-month global analysis. Environ Pollut 2021 Feb 15;271:116326 [FREE Full text] [doi: 10.1016/j.envpol.2020.116326] [Medline: 33412447]
18. About the alert system. Unite Against COVID-19. URL: https://covid19.govt.nz/alert-system/covid-19-alert-system/ [accessed 2020-04-12]
19. IHME COVID-19 health service utilization forecasting team, Murray CJL. Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries. medRxiv. Preprint posted online on April 26, 2020. [doi: 10.1101/2020.04.21.20074732]
20. Variation in government responses to COVID-19. Blavatnik School of Government. URL: https://www.bsg.ox.ac.uk/research/publications/variation-government-responses-covid-19 [accessed 2020-06-24]
21. Vernaccini L, Doherty B, Marin Ferrer M. INFORM Epidemic Risk Index: support collaborative risk assessment for health threats. JRC Publications Repository. 2018. URL: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/inform-epidemic-risk-index-support-collaborative-risk-assessment-health-threats [accessed 2020-04-12]
22. World Bank country and lending groups. World Bank Data Help Desk. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2020-04-12]
23. Community mobility reports. Google. URL: https://www.google.com/covid19/mobility/ [accessed 2020-05-17]
24. Moraga P. Small area disease risk estimation and visualization using R. R J 2018;10(1):495. [doi: 10.32614/rj-2018-036]
25. Islam N, Sharp SJ, Chowell G, Shabnam S, Kawachi I, Lacey B, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. BMJ 2020 Jul 15;370:m2743 [FREE Full text] [doi: 10.1136/bmj.m2743] [Medline: 32669358]
26. Allévius B, Höhle M. An unconditional space–time scan statistic for ZIP‐distributed data. Scand J Statist 2018 Jun 05;46(1):142-159 [FREE Full text] [doi: 10.1111/sjos.12341]
27. Brooks M, Kristensen K, Benthem K, Magnusson A, Berg C, Nielsen A, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J 2017;9(2):378. [doi: 10.32614/rj-2017-066]
28. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018 Sep;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
29. Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19 - studies needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/NEJMp2002125] [Medline: 32074416]

30. Payne C. COVID-19 in Africa. Nat Hum Behav 2020 May;4(5):436-437. [doi: 10.1038/s41562-020-0870-5] [Medline: 32246105]
31. Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, Imperial College COVID-19 Response Team, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020 Aug;584(7820):257-261. [doi: 10.1038/s41586-020-2405-7] [Medline: 32512579]
32. Lyu W, Wehby GL. Community use of face masks and COVID-19: evidence from a natural experiment of state mandates in the US. Health Aff (Millwood) 2020 Aug;39(8):1419-1425. [doi: 10.1377/hlthaff.2020.00818] [Medline: 32543923]
33. Block P, Hoffman M, Raabe IJ, Dowd JB, Rahal C, Kashyap R, et al. Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. Nat Hum Behav 2020 Jun;4(6):588-596. [doi: 10.1038/s41562-020-0898-6] [Medline: 32499576]
34. Shuchman M. Low- and middle-income countries face up to COVID-19. Nat Med 2020 Jul;26(7):986-988. [doi: 10.1038/d41591-020-00020-2] [Medline: 32439872]
35. Sehra ST, George M, Wiebe DJ, Fundin S, Baker JF. Cell phone activity in categories of places and associations with growth in cases of COVID-19 in the US. JAMA Intern Med 2020 Dec 01;180(12):1614-1620. [doi: 10.1001/jamainternmed.2020.4288] [Medline: 32865556]
36. Coronavirus disease (COVID-19): how is it transmitted? World Health Organization. URL: https://www.who.int/news-room/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted [accessed 2020-12-08]
37. Changruenngam S, Bicout DJ, Modchang C. How the individual human mobility spatio-temporally shapes the disease transmission dynamics. Sci Rep 2020 Jul 09;10(1):11325. [doi: 10.1038/s41598-020-68230-9] [Medline: 32647225]
38. Maxmen A. How poorer countries are scrambling to prevent a coronavirus disaster. Nature 2020 Apr;580(7802):173-174. [doi: 10.1038/d41586-020-00983-9] [Medline: 32242110]
39. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
40. Sehra ST, Kishfy LJ, Brodski A, George MD, Wiebe DJ, Baker JF. Association of cell phone location data and trends in COVID-19 infections during loosening of stay-at-home restrictions. J Travel Med 2020 Dec 23;27(8):taaa177 [FREE Full text] [doi: 10.1093/jtm/taaa177] [Medline: 32970146]
41. Parnell A, Goniewicz K, Khorram-Manesh A, Burkle FM, Al-Wathinani A, Hertelendy AJ. COVID-19 a health reform catalyst? —Analyzing single-payer options in the U.S.: considering economic values, recent proposals, and existing models from abroad. J Hosp Adm 2020 Aug 04;9(4):10. [doi: 10.5430/jha.v9n4p10]
42. Khorram-Manesh A, Carlström E, Hertelendy AJ, Goniewicz K, Casady CB, Burkle FM. Does the prosperity of a country play a role in COVID-19 outcomes? Disaster Med Public Health Prep 2020 Aug 12:1-10 [FREE Full text] [doi: 10.1017/dmp.2020.304] [Medline: 32782059]
43. Purnhagen KP, De Ruijter A, Flear ML, Hervey TK, Herwig A. More competences than you knew? The web of health competence for European Union action in response to the COVID-19 outbreak. Eur J Risk Regul 2020 Apr 16;11(2):297-306. [doi: 10.1017/err.2020.35]
44. Silver L, Johnson C. 1. Majorities in sub-Saharan Africa own mobile phones, but smartphone adoption is modest. Pew Research Center. 2018. URL: https://www.pewresearch.org/global/2018/10/09/majorities-in-sub-saharan-africa-own-mobile-phones-but-smartphone-adoption-is-modest/ [accessed 2020-06-24]

Abbreviations

INFORM: Index for Risk Management
LIC: low-income country
LMIC: low-income and middle-income country
NB: negative binomial
RR: relative risk
SIR: standardized incidence ratio
WHO: World Health Organization

Edited by G Eysenbach, C Basch; submitted 29.07.20; peer-reviewed by M Miyachi, J Li, I Mircheva, R De Carvalho, K Goniewicz, A Azzam; comments to author 17.11.20; revised version received 29.01.21; accepted 21.02.21; published 14.06.21.

Please cite as:
Tyrovolas S, Giné-Vázquez I, Fernández D, Morena M, Koyanagi A, Janko M, Haro JM, Lin Y, Lee P, Pan W, Panagiotakos D, Molassiotis A
Estimating the COVID-19 Spread Through Real-time Population Mobility Patterns: Surveillance in Low- and Middle-Income Countries
J Med Internet Res 2021;23(6):e22999
URL: https://www.jmir.org/2021/6/e22999
doi:10.2196/22999
PMID:33950850

Original Paper

Political Partisanship and Antiscience Attitudes in Online Discussions About COVID-19: Twitter Content Analysis

Ashwin Rao[1], BS; Fred Morstatter[1], PhD; Minda Hu[1], MS; Emily Chen[1], BS; Keith Burghardt[1], PhD; Emilio Ferrara[1], PhD; Kristina Lerman[1], PhD

[1]Information Sciences Institute, University of Southern California, Marina del Rey, CA, United States

Corresponding Author:
Ashwin Rao, BS
Information Sciences Institute
University of Southern California
4676 Admiralty Way
STE 1001
Marina del Rey, CA, 90292
United States
Phone: 1 213 505 0363
Email: mohanrao@usc.edu

Abstract

Background: The novel coronavirus pandemic continues to ravage communities across the United States. Opinion surveys identified the importance of political ideology in shaping perceptions of the pandemic and compliance with preventive measures.

Objective: The aim of this study was to measure political partisanship and antiscience attitudes in the discussions about the pandemic on social media, as well as their geographic and temporal distributions.

Methods: We analyzed a large set of tweets from Twitter related to the pandemic, collected between January and May 2020, and developed methods to classify the ideological alignment of users along the moderacy (hardline vs moderate), political (liberal vs conservative), and science (antiscience vs proscience) dimensions.

Results: We found a significant correlation in polarized views along the science and political dimensions. Moreover, politically moderate users were more aligned with proscience views, while hardline users were more aligned with antiscience views. Contrary to expectations, we did not find that polarization grew over time; instead, we saw increasing activity by moderate proscience users. We also show that antiscience conservatives in the United States tended to tweet from the southern and northwestern states, while antiscience moderates tended to tweet from the western states. The proportion of antiscience conservatives was found to correlate with COVID-19 cases.

Conclusions: Our findings shed light on the multidimensional nature of polarization and the feasibility of tracking polarized opinions about the pandemic across time and space through social media data.

(J Med Internet Res 2021;23(6):e26692) doi:10.2196/26692

KEYWORDS

COVID-19; Twitter; infodemiology; infodemic; infoveillance; multidimensional polarization; social media; social network

Introduction

Effective response to a health crisis requires society to forge a consensus on many levels: scientists and doctors have to learn about the disease and quickly and accurately communicate their research findings to others, public health professionals and policy experts have to translate the research into policies and regulations for the public to follow, and the public has to follow guidelines to reduce infection spread. However, the fast-moving COVID-19 pandemic has exposed our critical vulnerabilities at all these levels. Instead of orderly consensus-building, we have seen disagreement and controversy that exacerbated the toll of the disease. Research papers are rushed through the review process, with results sometimes being disputed or retracted [1], policy makers giving conflicting advice [2], and scientists and many in the public disagreeing on many issues, from the benefits of therapeutics [3] to the need for lockdowns and face-covering [4]. The conflicting viewpoints create conditions for polarization to color perceptions of the pandemic [5-8] and attitudes toward mitigation measures.

Surveys have identified a partisan gulf in the attitudes about COVID-19 and the costs and benefits of mitigation strategies, with the public's opinion polarized into sharply contrasting positions. According to a Pew Research Center report [9], political partisanship significantly affects perceptions of public health measures and might explain regional differences in the pandemic's toll in the United States [10]. Polarization has colored the messages of US political leaders about the pandemic [7] as well as discussions of ordinary social media users [8]. Coupled with a distrust of science and institutions, polarization can have a real human cost if it leads the public to minimize the benefits of face coverings or reject the COVID-19 vaccine when it becomes available. Dr Anthony Fauci, the nation's top infectious diseases expert, attributed many of the disease's 500,000 deaths (and counting) to political divisions in the country [11]. This further affirms the need to investigate the presence, and unravel the ill effects, of polarization in scientific and political discourse.

Current research measures polarization as divergence of opinions along the political dimension and its effect on other opinions, for example, discussion of scientific topics [12]. However, opinions on controversial issues are often correlated [13]; for example, those who support transgender rights also believe in marriage equality, and those who oppose lockdowns also resist universal face-covering. Inspired by this idea, we capture some of the complexity of polarization by projecting opinions in a multidimensional space, with different axes corresponding to different semantic dimensions. Once we identify the dimensions of polarization and define how to measure them, we can study the dynamics of polarized opinions, their interactions, and regional differences.

Our work analyzed tweets posted on Twitter related to the COVID-19 pandemic collected between January 21 and May 1, 2020 [5]. We studied polarization along three dimensions: political (liberal vs conservative), science (proscience vs antiscience), and moderacy (hardline vs moderate). User polarization along the science axis identifies whether users align with scientific and factual sources of information or whether they are characterized by mistrust of science and preference for pseudoscientific and conspiracy sources. A user's political ideology is defined in a 2D space. Working in tandem with the political axis, the moderacy dimension recognizes the intensity of partisanship from hardline to moderate. For the hardliners identified along the moderacy dimension, we leveraged the political axis to identify their partisanship as liberal or conservative.

Cinelli et al [14] and Weld et al [15] showed that sharing of URLs annotated by Media Bias/Fact Check is a reliable proxy of one's political polarity. Inspired by the findings and conclusions made in these works, we used media sources that have been classified by nonpartisan sites along these dimensions to define the poles of each dimension of polarization. These media sources include both mainstream news and a large variety of other sources, such as government agencies, nongovernmental organizations, crowdsourced content, and alternative medicine news and health sites. Users were given a score reflecting how often they shared information from each set of polarized sources. These users served as training data to train machine learning algorithms to classify remaining users along the multiple dimensions of polarization based on the content of their posts. Inferring the polarization of users discussing COVID-19 allowed us to study the relationships between polarized ideologies and their temporal and geographic distributions. We showed that political and science dimensions were highly correlated and that politically hardline users were more likely to be antiscience, while politically moderate users were more often proscience. We also identified regions of the United States and time points where the different ideological subgroups were comparably more active and we identified their topics of conversation. We found that areas of heightened antiscience activity corresponded to US states with large COVID-19 outbreaks. Our work, therefore, provides insights into potential reasons for geographic heterogeneity of outbreak intensity.

The contributions of this work are as follows:

- We described a framework to infer the multidimensional polarization of social media users, allowing us to track political partisanship and attitudes toward science at scale.
- We showed that political and science dimensions were highly correlated, with hardline right and antiscience attitudes closely aligned.
- We studied the geographical distribution of polarized opinions and found that regional differences can correlate with the pandemic's toll.

As the amount of COVID-19 information explodes, we need the ability to proactively identify emerging areas of polarization and controversy. Early identification could lead to more effective interventions to reduce polarization and also improve the efficacy of disease mitigation strategies. Vaccine hesitancy was shown in past research to be associated with antiscience attitudes [16]; therefore, our approach may help identify regions of the country that will be more resistant to COVID-19 vaccination. This may better prepare public health workers to target their messages.

Methods

Here, we describe the data and methods we used for measuring polarization and also inferred it from text and online interactions.

Data Set

In this study, we used a public data set of COVID-19 tweets from Twitter [5]. This data set comprises 115 million tweets from users across the globe, collected over a period of 101 days from January 21 to May 1, 2020. These tweets contain at least one keyword from a predetermined set of COVID-19–related keywords (eg, coronavirus, pandemic, and Wuhan).

Fewer than 1% of the tweets in the original corpus have geographic coordinates associated with them. We specifically focused on tweets from users located in the United States, at state-level granularity, based on geolocated tweets and fuzzy matching of user profile text [8]. Specifically, we used a fuzzy text matching algorithm to detect state names and abbreviations, as well as names of populous cities. The user profile text extracted from the *description* attribute of the *user* object was passed on to the *loc_to_state* function of the georeferencing code [17] to extract the user's location at the state level. A manual review of this approach found it to be effective in identifying the user's home state. This methodology provided location information for 65% of users in the data set. The georeferenced data set consisted of 27 million tweets posted by 2.4 million users over the entire time period.

Measuring Polarization Using Domain Scores

We characterized individual attitudes along three dimensions of polarization. The *political* dimension, the standard dimension for characterizing partisanship, captured the difference between *left* (*liberal*) and *right* (*conservative*) stances for users with strong hardline political opinions. The *science* dimension captured an individual's acceptance of evidence-based *proscience* views or the propensity to hold *antiscience* views. People believing and promoting conspiracies, especially health-related and pseudoscientific conspiracies, were often grouped in the antiscience camp. Finally, the *moderacy* dimension described the intensity of partisanship, from *moderate* or nonpartisan opinions to politically *hardline* opinions.

We inferred polarized attitudes of users from the content of their posts. While previous work [18] inferred polarization from user hashtags, we instead relied on user-tweeted URLs. The key idea that motivated our approach is that online social networks tend to be ideologically similar, with users more closely linked (eg, through follower relationships) to others who share their beliefs [19,20]. While we did not have follow links in our data, we used URLs as evidence [21] of a homophilic link. We extended this approach beyond political ideology [22] to label other dimensions of polarization. Specifically, we used a curated list of information sources, whose partisan leanings were classified by neutral websites, to infer the polarization of Twitter users at scale. We used lists compiled by Media Bias/Fact Check, AllSides, and NewsGuard, which tracks coronavirus misinformation (see data folder at GitHub [23]). Table 1 lists exemplar domains, hereinafter referred to as pay-level domains (PLDs), in each category. PLDs listed under *conspiracy* and *questionable sources* were mapped to our antiscience category. For the *moderacy* axis, we considered the union of left and right PLDs as a proxy for the *hardline* category, while the union of least-biased, left-moderate, and right-moderate PLDs formed the proxy *moderate* category.

We quantified a user's position along the dimensions of polarization by tracking the number of links to curated PLDs the user shared. Specifically, we extracted PLDs that were shared by users in the data set and filtered for relevant PLDs that were present in our curated lists (Table 1). This gave us a set of 136,000 users who shared *science* PLDs, 169,000 users who shared *political* PLDs, and 234,000 users who shared PLDs along the *moderacy* dimension. There was a wide distribution in the number of tweets, and therefore PLDs, shared between users (Figure S1 in Multimedia Appendix 1), with some users tweeting many PLDs and many users tweeting one or none. We, therefore, filtered out users who shared fewer than *three* relevant PLDs in each dimension (ie, fewer than three in the science dimension, fewer than three in the partisan dimension, and fewer than three in the moderacy dimension), which resulted in 18,700 users. For each user, we computed a *domain score* δ along each of the three dimensions, as the average of mapped domain values of a dimension:

where δ_i is the domain score of $user_i$ and $D_{i,d}$ represents the set of PLDs shared by $user_i$ relevant to dimension d.

Table 1. Curated information and news pay-level domains (PLDs) with their polarization.

Dimension and polarization dimension	PLDs, n	Examples of PLDs
Science[a]		
Proscience (+1)	150+	cdc.gov, who.int, thelancet.com, mayoclinic.org, nature.com, and newscientist.com
Antiscience (−1)	450+	911truth.org, althealth-works.com, naturalcures.com, shoebat.com, and prison-planet.com
Political[b]		
Liberal (−1)	300+	democracynow.org, huffingtonpost.com, newyorker.com, occupy.com, and rawstory.com
Conservative (+1)	250+	nationalreview.com, newsmax.com, oann.com, theepochtimes.com, and bluelivesmatter.blue
Moderacy[c]		
Moderate (+1)	400+	ballotpedia.org, c-span.org, hbr.org, wikipedia.org, weforum.org, snopes.com, and reuters.com
Hardline (−1)	500+	gopusa.com, cnn.com, democracynow.org, huffingtonpost.com, oann.com, and theepochtimes.com

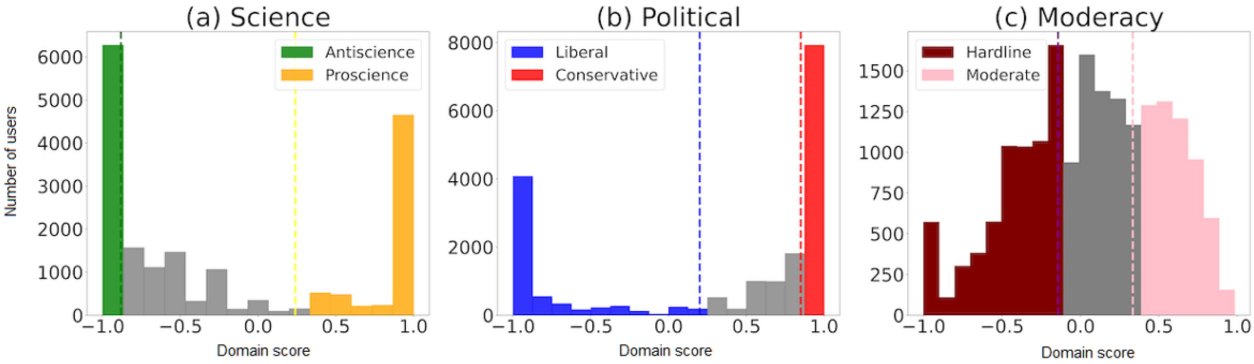
[a] Proscience PLDs are mapped to +1 along the science axis, while antiscience PLDs are mapped to −1.

[b] Along the political axis, liberal PLDs are mapped to −1, while conservative PLDs are mapped to +1.

[c] Along the moderacy axis, hardline PLDs are mapped to −1, while moderate PLDs are mapped to +1.

Figure 1 shows the distribution of domain scores for users who shared links to information sources across all dimensions. The distributions were peaked at their extreme values, showing more users sharing information from antiscience than proscience PLDs and more conservative than liberal PLDs. In Figure S2 in Multimedia Appendix 1, we show that these extremes were robust to how we filtered users and were, therefore, not a product of, for example, sharing a single link.

Figure 1. The distribution of domain scores along science, political, and moderacy dimensions. (a) The vertical lines at 0.42 and −1 mark the top and bottom 30% cutoffs of distribution along the science dimension, which are binned as proscience (+1) and antiscience (−1), respectively. (b) The vertical lines at 1 and −0.33 mark the top and bottom 30% cutoffs of distribution along the political dimension, which are binned as conservative (+1) and liberal (−1), respectively. (c) The vertical lines at 0.38 and −0.18 mark the top and bottom 30% cutoffs of distribution along the moderacy dimension, which are binned as moderate (+1) and hardline (−1), respectively.



For network-level analysis, we then built a web scraper that mapped PLDs to their respective Twitter handles. The scraper initiated a simple Google query of the form "*Domain Name Twitter Handle*." This tool relied on the search engine to rank results based on relevance and picked out the title of the first result containing the substring "|*Twitter*." This substring was of the form "*Account Name* (*@handle*) | *Twitter*," which was parsed to retrieve the domain's corresponding handle. We manually verified the mapped PLDs. The mapped dimension-wise PLDs are available on our GitHub repository under the data folder.

Recall that along each of the three dimensions, we mapped the dimension's constituent domain names to their respective Twitter handles. The mapped Twitter handles formed our seed sets for semisupervised learning at the network level. Each dimension's seed set comprised key-value pairs of Twitter handles and their corresponding orientation along the dimension.

Table 2 illustrates the number of seeds along each polarization axis.

To investigate the presence of bias stemming from an uneven distribution of PLDs along each ideological dimension's polarized ends, we sampled an equal number of PLDs along each of the dimension's two polarities. More specifically, we performed random downsampling of the majority ideological polarity along each dimension. Upon ensuring that each dimension's polarized ends were now represented by an equal number of PLDs, we calculated domain scores for each user along the ideological dimensions (Figure S3 in Multimedia Appendix 1). Leveraging these domain scores, we then rebuilt prediction models. We found that the performance of this modified procedure did not differ significantly from our results (see Table S1 in Multimedia Appendix 1 for more details). This robustness check demonstrated the versatility of our approach to differences in the sampling of PLDs along each dimension.

Table 2. Description of the retweet network.

Dimension and polarization	Seeds[a], n (%)
Science (n=158)	
Proscience	81 (51.3)
Antiscience	77 (48.7)
Political (n=195)	
Liberal	96 (49.2)
Conservative	99 (50.8)
Moderacy (n=558)	
Hardline	195 (34.9)
Moderate	363 (65.1)

[a] Number of seed handles along each polarization axis for initial node assignment in the label propagation algorithm.

Inferring Polarization

Overview

Using domain scores, we were able to quantify the polarization of just a small fraction (18,700/2,400,000, 0.78%) of users who generated PLDs in the data set. In this section, we describe how we leveraged this data to infer the polarization of the remaining users in our data set. In the Results section, we compare the performance of these inference methods. Two methods, label propagation algorithm (LPA) and latent Dirichlet allocation (LDA), act as baselines against our state-of-the-art text embedding method. Our study focused on investigating content generated by users over the entire period rather than at the noisier tweet level. Investigating a user's content, tweet by tweet, may or may not provide sufficient information to gauge their ideological polarity, whereas analyzing all tweets generated by a user over time would facilitate this.

We classified users according to the binned domain scores along each dimension. We found that classification worked better than regression in this data set. We binned domain scores by thresholding the distribution into two classes along each dimension, as shown in Figure 1. Using other threshold values to bin the domain score distribution into two classes did not qualitatively change results (Multimedia Appendix 1). Additionally, we released a GitHub repository [23] for readers to reproduce this work upon careful rehydration of tweet data, instructions for which have also been provided in the repository.

Label Propagation Algorithm

LPA was used in the past to label user ideology based on the ideology of accounts the user retweets (eg, see Badawy et al [22]). The idea behind label propagation is that people prefer to connect to, and retweet content posted by, others who share their opinions [24,25]. This gives us an opportunity to leverage topological information from the retweet network to infer users' propensity to orient themselves along ideological dimensions.

The geocoded Twitter data set provides fields named *screen_name* and *rt_user*, which allowed us to identify the user being retweeted and the user retweeting, respectively. To this end, we built a network from 9.8 million retweet interactions between 1.9 million users sourced from the data set. In the retweet network, an edge runs from *A* to *B* if user *A* retweets user *B*. Descriptive statistics of the retweet network are shown in Table 3. We then used the semisupervised greedy learning algorithm (ie, the LPA) to identify clusters in the retweet network.

LPA, as proposed by Raghavan et al [26], is a widely used near-linear time node classification algorithm. This greedy learning method started off with a small set of labeled nodes also known as seeds, with the remaining nodes assigned labels at random. The number of seeds for each polarization dimension is shown in Table 2. The algorithm then iteratively updated the labels of nonseed nodes to the majority label of their neighbors, with ties broken at random, until converging to an equilibrium where the labels no longer changed. However, owing to stochasticity of tie-breaking, a certain amount of randomness crept into the results produced by this algorithm. As the result, LPA tended to generate slightly differing classifications of user polarization for the same network each time it was run. To address the stochasticity, we ran the LPA in 5-fold cross-validation and averaged the results.

Table 3. Statistics of the network.

Statistic	Value, n
Nodes	1,857,028
Maximum in-degree	39,149
Maximum out-degree	1450
Retweets	9,788,251
Unique retweets	7,745,533
Size of the strongly connected component	1,818,657

Latent Dirichlet Allocation

We used LDA [27] to identify topics, or groups of hashtags, and represented users as vectors in this topic space. We considered the set of *all* hashtags in the COVID-19 data set generated by a user over time as a document representing that user—after ignoring hashtags used by fewer than 10 users or more than 75% of the users—leaving 25,200 hashtags. The choice of 75% was arbitrary, but a hashtag that appeared at a lower threshold (eg, within roughly 50% of the users) could be highly prevalent in one domain and not another. We used a more lenient threshold to avoid this issue. We also used 20 topics, as that gave a higher coherence score. Given the enormity of the geocoded Twitter data set we leveraged in this study, conducting LDA experiments to validate these thresholds proved to be computationally prohibitive and it was unlikely that tuning would have achieved significantly better results than the one seen in this study.

We used the document-topic affinity matrix generated by LDA to represent users. An *affinity vector* was composed of 20 likelihood scores corresponding to 20 topics, adding up to 1, with each score indicating the probability of the corresponding topic being a suitable representation for the set of hashtags generated by the user. Using these affinity vectors, we generated feature vector matrices for each of the three dimensions of interest. In doing so, we were able to represent over 900,000 users who used some hashtag in their tweets with a dense vector of length 20.

Text Embedding Using fastText

Previous methods—see Conover et al [28]—classified a user's political polarization based on the text of their tweets by generating term frequency–inverse document frequency–weighted unigram vectors for each user. However, the advent of more powerful text-embedding techniques [29-31] allowed us to generate sentence-embedding vectors to better represent content.

We grouped the tweets generated by each of the 2.4 million users from January to May 2020. More specifically, we collected all COVID-19–related tweets generated by a user in this time period and concatenated them to form a text document for each user. After preprocessing the 2.4 million documents to lowercase and removing hashtags, URLs, mentions, handles, and stop words, we used the fastText sentence-embedding model pretrained on Twitter data to generate tweet embeddings for each user. Preprocessing of tweets was performed by leveraging the regular expression (*re*) package in Python, version 3.7 (Python Software Foundation); the Natural Language Toolkit; and the *Gensim* natural language processing library. The *Sent2vec* Python package [32] provided us with a Python interface to quickly leverage the pretrained model and generate 700-dimension feature vectors representing each user's discourse.

Results

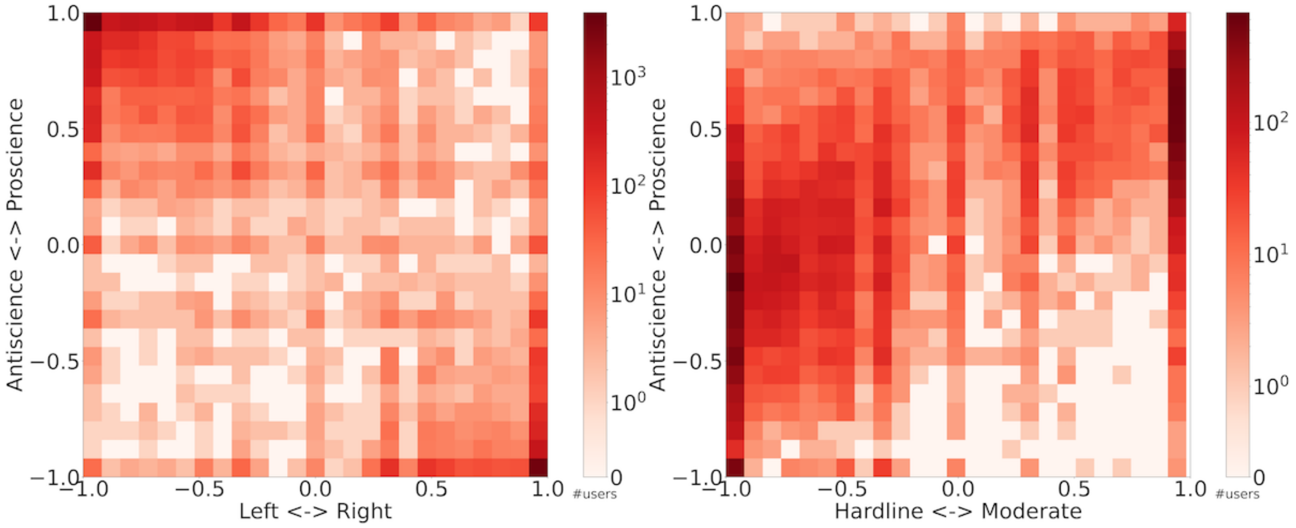
Overview

First, we visualized the domain scores of the 18,700 users, showing the relationship between the science, moderacy, and political dimensions. Then we compared the performance of algorithms for classifying users along the three dimensions of polarization, using domain scores as ground truth data. We used the inferred scores to study the dynamics and spatial distribution of polarized opinions of users engaged in online discussions about COVID-19.

Visualizing Polarization

Figure 2 shows the relationship between dimensions of polarization, leveraging domain scores of 18,700 users who shared information from curated PLDs. The heat map shows the density of users with specific domain scores. Large numbers of users are aligned with proscience-left extreme (top-left corner) or antiscience-right extreme (bottom-right corner), with lower densities along the diagonal between these extremes (Figure 2, left-hand side). This illustrates the strong correlation between political partisanship and scientific polarization, thereby highlighting the influence of pernicious political divisions on evidence-based discourse during the pandemic, with conservatives being more likely to share antiscience information than proscience sources. The heat map on the right-hand side in Figure 2 highlights the interplay between the science and moderacy axes. The white region in the bottom-right corner shows there are few antiscience users who are politically moderate, thus demonstrating an asymmetry in these ideologies. The shading also highlights a higher density of proscience users identifying as politically moderate. These results are robust to how data are filtered, as shown in Figure S4 in Multimedia Appendix 1.

Figure 2. Polarization of COVID-19 tweets. On the left is the heat map of polarization (domain scores) along the science-partisanship dimensions. On the right is the heat map of polarization (domain scores) along the science-moderacy dimensions. Each bin within the heat map represents the number of users with domain scores falling within that bin.



Classifying Polarization

To run the LPA, we started from a set of labeled seeds: Twitter handles corresponding to PLDs categorized along the dimensions of interest (Tables 1 and 2). We reserved some of the seeds along each dimension for testing LPA predictions and reported accuracy of 5-fold cross-validation.

For content-based approaches, we used binned domain scores of 18,700 users as ground truth data to train logistic regression models to classify user polarization along the three dimensions. We represented each user as a vector of features generated by different content-based approaches: topic vectors for LDA and sentence embeddings for the fastText approach. We reserved a subset of users for testing performance.

Table 4 compares the performance of polarization classification methods. LPA worked well when it tried to identify user alignment along the political and science dimensions. However, it failed to capture the subtler distinctions along the moderacy axis. Training was further hampered by the low number of retweet interactions with moderate PLDs in comparison to hardline ones. Of the 1.8 million retweet interactions, only 250,000 involved some moderate seed nodes, whereas over 1 million interactions involved some hardline seed nodes. Moreover, poor classification performance with LPA revealed an important pattern: that moderates surrounded themselves with diverse opinions and, thus, a clear distinction could not be made by observing who they retweeted.

LDA modeling on hashtags allowed us to generate reduced-dimension, dense feature vectors for over 900,000 users who used hashtags in their tweets. This representation allowed us to design better learning models that significantly outperformed the LPA model.

A logistic regression model trained on fastText outperformed all other models described in this study. Coupled with fastText's ability to better handle out-of-vocabulary terms, the model's access to finer levels of detail at the tweet-text level, culminated in it being able to better predict dimensions of polarization. Given the model's superior performance across all three dimensions, we leveraged its predictions in subsequent analyses. We classified users along the three polarization dimensions. However, since the definition of the hardline extreme of the moderacy dimension overlapped with the political dimension, we needed to report only six ideological groups, rather than all eight combinations.

Table 4. Performance of polarization classification.[a]

Method and dimension	Data set size, n	Accuracy, %	Precision, %	Recall, %	F1 score, %
Label propagation algorithm					
Science	158	92.6	*100* [b]	80	88.9
Political	195	92.3	86.9	*100*	93.0
Moderacy	1205	20.1	72	1.4	2.74
Latent Dirichlet allocation					
Science	9983	92.2	91.6	92.4	91.9
Political	11,020	93.5	95.1	93.3	94.2
Moderacy	9565	86.4	85.6	85.0	85.4
fastText					
Science	11,202	*93.8*	93.9	*93.7*	*93.8*
Political	12,425	*95.1*	*96.5*	94.6	*95.5*
Moderacy	11,197	*90.2*	*90.1*	*90.5*	*90.2*

[a] Results compare classification performance of the label propagation algorithm and content-based methods, including topic modeling (latent Dirichlet allocation) and full-text embedding (fastText). Results are averages of 5-fold cross-validation. Data set sizes are the number of users in model validation data sets and are composed of users with strong polarization scores (top or bottom 30% as defined previously) in the filtered 18,700-user data set.

[b] Values in italics indicate the best-performing models.

Discussion

Dynamics of Polarization

Research shows that opinions of Twitter users about controversial topics do not change over time [33]. To investigate whether user alignments along the three polarization dimensions changed over time, we grouped tweets by time into seven biweekly intervals: January 21 to 31, 2020; February 1 to 15, 2020; February 16 to 29, 2020; March 1 to 16, 2020; March 17 to 31, 2020; April 1 to 15, 2020; and April 16 to May 1, 2020. There were 3000 users who tweeted consistently in all seven biweekly intervals. For each of the N users, we computed cumulative domain scores along science, political, and moderacy dimensions for all time intervals t and computed the average absolute change in domain score from biweekly period $t-1$ along each dimension given by the following:

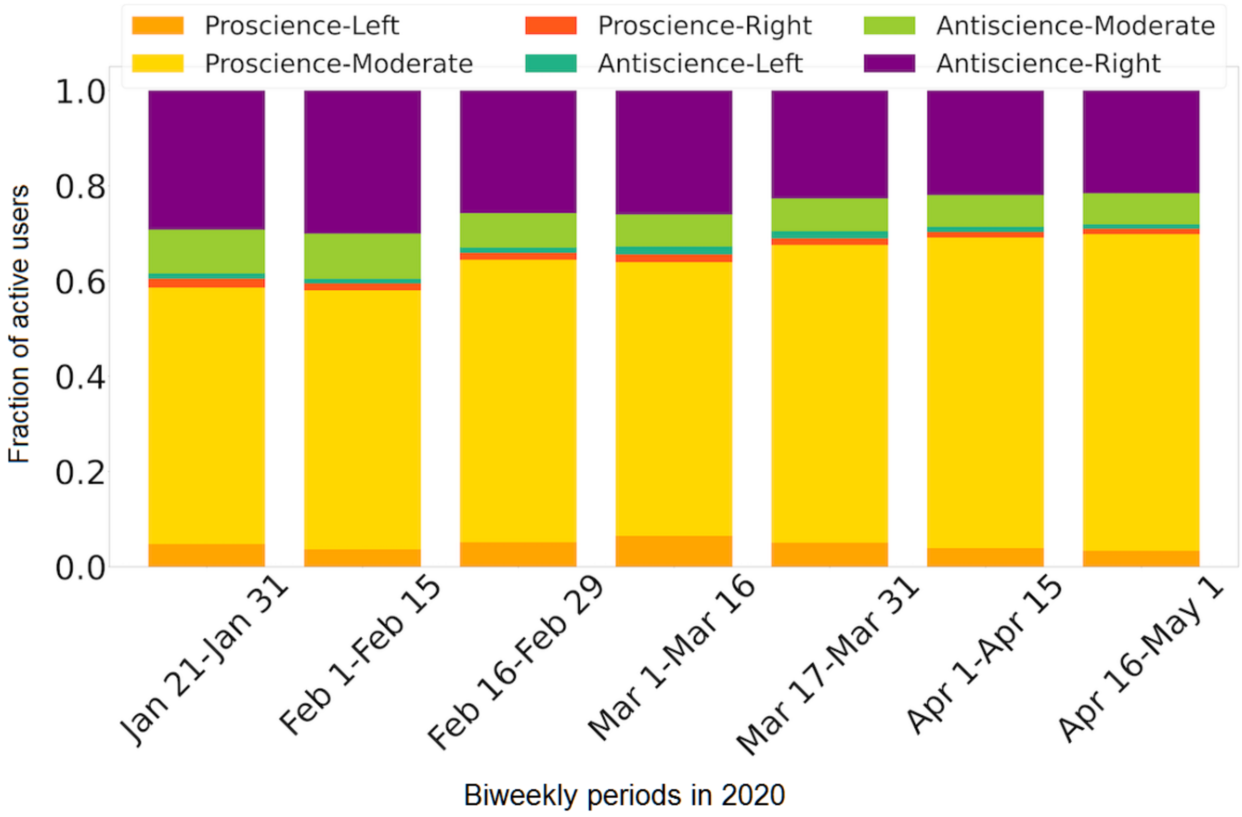
where $\delta_{i,t}$ represents the domain score for a user i in biweekly period t. The small values of in Table 5 confirm that user alignments do not change significantly over time.

Table 5. Average absolute change in domain score along consecutive biweekly intervals.

Dimension	Average absolute change () per biweekly interval numbers					
	2,1	3,2	4,3	5,4	6,5	7,6
Political	0.09	0.05	0.03	0.02	0.03	0.02
Science	0.13	0.07	0.04	0.02	0.02	0.02
Moderacy	0.21	0.11	0.07	0.04	0.04	0.03

Although each individual's alignments did not change, the number of users within each ideological group did change over time. User alignments did not change; therefore, we leveraged polarization classification results to show biweekly fractions of active users per ideological category. Figure 3 shows the composition of active users in all categories. As time progressed, we could clearly see the growth in the proscience-moderate category accompanied by a corresponding decline in antiscience-right users. This was consistently found over a variety of data filters, as seen in Figure S5 in Multimedia Appendix 1.

Figure 3. Fraction of active users per ideological group in biweekly periods. For completeness, this plot shows all users in the data set and not the filtered 18,700 users.



Topics of Polarization

To better understand what each of the six groups tweeted about, we collected the 50 most frequent hashtags used by each group, after removing hashtags common to all six groups. Figure 4 shows the word clouds of the most common hashtags within each group, sized by the frequency of their occurrence. Most striking was the use of topics related to conspiracy theories, such as *#qanon* and *#wwg1wga* by the antiscience-right group, along with politically charged references to the *#ccpvirus* and *#chinavirus*. This group also used hashtags related to former US President Donald Trump’s re-election campaign, showing the hyper-partisan nature of COVID-19 discussions. Another partisan issue appeared to be *#hydroxychloroquine*, a drug promoted by Donald Trump. It showed up in both proscience-right and antiscience-right groups but was not discussed by other groups. Overall, these intuitive results highlight the overall accuracy of our polarization inference model.

Figure 4. Topics of discussion within the six ideological groups. The top row (from left to right) illustrates topics for proscience-left, proscience-moderate, and proscience-right groups. The bottom row (from left to right) illustrates topics for antiscience-left, antiscience-moderate, and antiscience-right groups.

(a) Proscience-Left (b) Proscience-Moderate (c) Proscience-Right

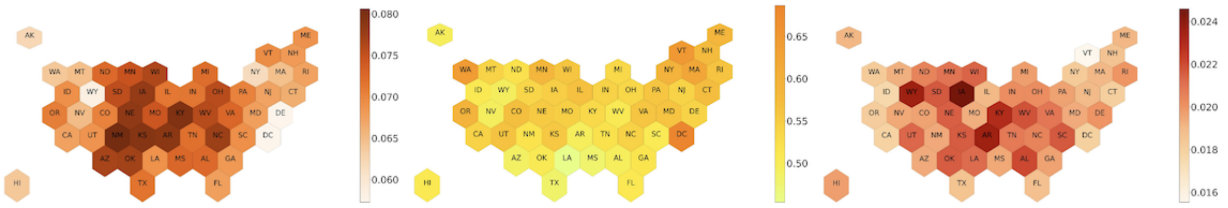
(d) Antiscience-Left (e) Antiscience-Moderate (f) Antiscience-Right

The polarized nature of the discussions could be seen in the users of the hashtags *#trumppandemic* and *#trumpvirus* by the left and proscience groups. However, in contrast to antiscience groups, proscience groups talked about COVID-19 mitigation strategies, using hashtags such as *#stayhomesavelives*, *#staysafe*, and *#flattenthecurve*.

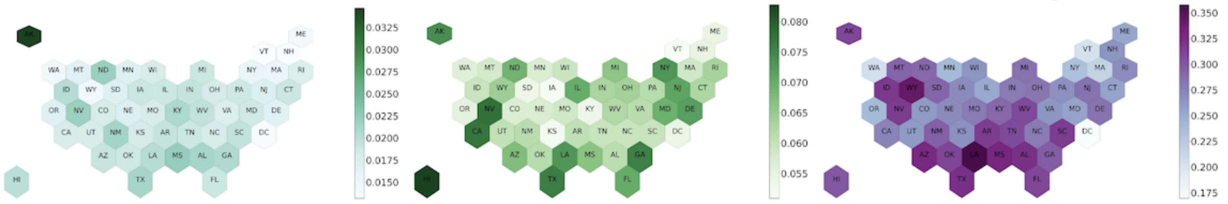
Geography of Polarization

Responses to the coronavirus pandemic in the United States have varied greatly by state. While the governors of New York, California, Ohio, and Washington reacted early by ordering lockdowns, the governors of Florida and Mississippi have downplayed the gravity of the situation for a longer time. To explore the geographical variation in ideological alignments, we grouped users by the state from which they tweeted and computed the fraction of their respective state's Twitter users belonging to an ideological group. We then generated geo-plots, shown in Figure 5, to highlight the ideological composition of each state.

Figure 5. Fraction of US states' Twitter users per ideological category. Plots (a) to (c) (top row, left to right) show the fraction of states' Twitter users who were classified as proscience-left, proscience-moderate, and proscience-right, respectively. Plots (d) to (f) (bottom row, left to right) show the fraction of states' Twitter users who were classified as antiscience-left, antiscience-moderate, and antiscience-right, respectively. The vertical bars next to the maps indicate the fraction of Twitter users in the state belonging to the ideological group. Two-letter abbreviations are used for each state.



(a) Proscience-Left (b) Proscience-Moderate (c) Proscience-Right



(d) Antiscience-Left (e) Antiscience-Moderate (f) Antiscience-Right

We saw a higher composition of proscience-moderates, as seen in Figure 5 (b), in Washington, Oregon, DC, and Vermont. As expected, these states had a lower fraction of antiscience users, as can be seen from Figure 5 (d), (e), and (f). Governors of these states were quick to enforce lockdowns and spread pandemic awareness among the general public.

Over the course of the pandemic, we have seen the strong opposition to masking mandates and closing down of businesses in California, Nevada, Hawaii, Georgia, and Texas. These antiscience sentiments are reflected in Figure 5 (e), which shows that these states had a comparatively higher proportion of their Twitter users in the antiscience-moderate ideology group.

Southern states—South Carolina, Mississippi, Louisiana, Texas, and Arizona—and northwestern states—Wyoming, North Dakota, South Dakota, and Montana—have experienced COVID-19 surges, with southern states becoming overwhelmed during the summer of 2020 and northwestern states becoming overwhelmed in the fall of 2020 (Figure S6 in Multimedia Appendix 1 shows the cumulative COVID-19 cases per state). The political and religious leaders in these states have also consistently downplayed the pandemic and resisted mitigation strategies. Our results are consistent with this, showing that these states also had more conservative Twitter users who mistrust science, as manifested by sharing information from antiscience sources. The antiscience attitudes in these states may also spell trouble for vaccination plans. The statistically significant positive correlation (Figure S7 in Multimedia Appendix 1) between state-wise cumulative COVID-19 case counts and antiscience-right users, as well as the negative correlation between the former and proscience-moderate users, affirms the significance of scientific beliefs in mitigating the spread of virus.

Limitations and Future Directions

Our novel approach to identify ideological alignments of users on Twitter comes with certain limitations. Akin to other studies involving Twitter data, our study worked under the caveat that the behaviors of the subset of users being considered in our data set may not be representative of population behavior. The use of geolocation techniques and subsequent consideration of users with a geolocation could introduce certain biases, which necessitate further investigation.

Thresholds that were used in our LDA analysis of user hashtags have been set intuitively due to LDA's prohibitive computation needs when dealing with over 900,000 hashtags. It is unlikely that we would have observed significant improvements in classification results with different thresholds. However, we encourage readers to investigate this further.

Additionally, the seed sets (Table 2) employed for our label propagation experiments may have had room for bias, as not all PLDs collected had a corresponding Twitter account. The cross-section of PLDs that have a Twitter account could be biased by political orientation, age group that the PLD caters to, etc. Investigation of bias stemming from this is a promising prospect for future work. Furthermore, our analyses worked under the assumption that media bias ratings provided by Media Bias/Fact Check accurately exhibited ideological biases of media sources. Leveraging these ratings, we assumed that generating tweets consisting of PLDs was an expression or reflection of one's ideological polarity. Future studies can build on these assumptions, and interesting avenues can be explored by incorporating other indicators of user polarity.

Verification of agreement or disagreement of user viewpoint and content in PLDs being shared was not in the purview of this study, and we encourage our readers to explore these avenues in future research. Furthermore, although we showed good performance on classifying polarized opinions, additional work is required to infer finer-grained opinions. Namely, by predicting fine-grained polarization among users, we could better infer, for example, network effects, such as whether users prefer to interact with more polarized neighbors, which may adversely impact provaccine mitigation strategies. Moreover, longer-term trends need to be explored in order to better understand how opinions change dynamically. This will better test whether social influence or selective formations of ties are the drivers of echo chambers and polarization. Finally, there is a need to explore polarization across countries to understand how different societies and governments are able to address polarization and how these polarized dimensions relate to one another across the world.

Conclusions

Our analysis of a large corpus of online discussions about COVID-19 confirms and extends the findings of opinion polls and surveys [9]: opinions about COVID-19 are strongly polarized along partisan lines. Political polarization strongly interacts with attitudes toward science: conservatives are more likely to share antiscience information related to COVID-19, while liberal and more moderate users are more likely to share information from proscience sources. On the positive side, we found that the number of proscience, politically moderate users dwarfed other ideological groups, especially antiscience groups. This is reassuring from the public health point of view, suggesting that a plurality of people are ready to accept scientific evidence and trust scientists to lead the way out of the pandemic. The geographical analysis of polarization identified regions of the country, particularly in the south and the west where antiscience attitudes are more common, that correlate to areas with particularly high COVID-19 cases, as seen in Figure S6 in Multimedia Appendix 1. Messaging strategies should be tailored in these regions to communicate with science skeptics. Overall, we found that analysis of tweets, while less representative than surveys, offers inexpensive, fine-grained, and real-time analysis of polarization and partisanship.

Acknowledgments

This research was sponsored, in part, by the Air Force Office for Scientific Research under contract FA9550-17-1-0327 and by DARPA (Defense Advanced Research Projects Agency) under contract W911NF-18-C-0011.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 706 KB - jmir_v23i6e26692_app1.docx]

References

1. Retracted coronavirus (COVID-19) papers. Retraction Watch. URL: https://retractionwatch.com/retracted-coronavirus-covid-19-papers/ [accessed 2021-05-28]
2. Schünemann HJ, Akl EA, Chou R, Chu DK, Loeb M, Lotfi T, et al. Use of facemasks during the COVID-19 pandemic. Lancet Respir Med 2020 Oct;8(10):954-955. [doi: 10.1016/s2213-2600(20)30352-0]
3. Zou L, Dai L, Zhang X, Zhang Z, Zhang Z. Hydroxychloroquine and chloroquine: A potential and controversial treatment for COVID-19. Arch Pharm Res 2020 Aug;43(8):765-772 [FREE Full text] [doi: 10.1007/s12272-020-01258-7] [Medline: 32740801]
4. Pleyers G. The pandemic is a battlefield. Social movements in the COVID-19 lockdown. J Civ Soc 2020 Aug 06;16(4):295-312 [FREE Full text] [doi: 10.1080/17448689.2020.1794398]
5. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus Twitter data set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
6. Druckman JN, Klar S, Krupnikov Y, Levendusky M, Ryan JB. Affective polarization, local contexts and public opinion in America. Nat Hum Behav 2021 Jan;5(1):28-38. [doi: 10.1038/s41562-020-01012-5] [Medline: 33230283]
7. Green J, Edgerton J, Naftel D, Shoub K, Cranmer SJ. Elusive consensus: Polarization in elite communication on the COVID-19 pandemic. Sci Adv 2020 Jun 24;6(28):eabc2717. [doi: 10.1126/sciadv.abc2717]
8. Jiang J, Chen E, Lerman K, Ferrara E. Political polarization drives online conversations about COVID-19 in the United States. Hum Behav Emerg Technol 2020 Jun 18:1 [FREE Full text] [doi: 10.1002/hbe2.202] [Medline: 32838229]
9. Funk C, Tyson A. Partisan differences over the pandemic response are growing. Pew Research Center. Washington, DC: Pew Research Center; 2020 Jun 03. URL: https://www.pewresearch.org/science/2020/06/03/partisan-differences-over-the-pandemic-response-are-growing/ [accessed 2021-06-07]
10. Gollwitzer A, Martel C, Brady WJ, Pärnamets P, Freedman IG, Knowles ED, et al. Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. Nat Hum Behav 2020 Nov;4(11):1186-1197. [doi: 10.1038/s41562-020-00977-7] [Medline: 33139897]
11. Steenhuysen J. Fauci says US political divisions contributed to 500,000 dead from COVID-19. Reuters. 2021 Feb 22. URL: https://www.reuters.com/article/us-health-coronavirus-fauci/fauci-says-u-s-political-divisions-contributed-to-500000-dead-from-covid-19-idUSKBN2AM2O9 [accessed 2021-06-07]

12. Bessi A, Zollo F, Del Vicario M, Puliga M, Scala A, Caldarelli G, et al. Users polarization on Facebook and Youtube. PLoS One 2016;11(8):e0159641 [FREE Full text] [doi: 10.1371/journal.pone.0159641] [Medline: 27551783]
13. Baumann F, Lorenz-Spreen P, Sokolov IM, Starnini M. Modeling echo chambers and polarization dynamics in social networks. Phys Rev Lett 2020 Jan 27;124(4):048301-1-048301-6. [doi: 10.1103/physrevlett.124.048301]
14. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):1-8 [FREE Full text] [doi: 10.1073/pnas.2023301118] [Medline: 33622786]
15. Galen W, Maria G, Tim A. Political bias and factualness in news sharing across more than 100,000 online communities. ArXiv. Preprint posted online on February 17, 2021. [FREE Full text]
16. Carrieri V, Madio L, Principe F. Vaccine hesitancy and (fake) news: Quasi‐experimental evidence from Italy. Health Econ 2019 Aug 20;28(11):1377-1382. [doi: 10.1002/hec.3937]
17. twitter-locations-us-state. GitHub. 2020. URL: https://github.com/julie-jiang/twitter-locations-us-state [accessed 2021-05-28]
18. Conover MD, Gonçalves B, Flammini A, Menczer F. Partisan asymmetries in online political activity. EPJ Data Sci 2012 Jun 18;1(1):1-19. [doi: 10.1140/epjds6]
19. Bakshy E, Messing S, Adamic LA. Political science. Exposure to ideologically diverse news and opinion on Facebook. Science 2015 Jun 05;348(6239):1130-1132. [doi: 10.1126/science.aaa1160] [Medline: 25953820]
20. Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R. Tweeting from left to right: Is online political communication more than an echo chamber? Psychol Sci 2015 Oct;26(10):1531-1542. [doi: 10.1177/0956797615594620] [Medline: 26297377]
21. Adamic LA, Glance N. The political blogosphere and the 2004 US election: Divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery. 2005 Presented at: 3rd International Workshop on Link Discovery; August 21, 2005; Chicago, IL p. 36-43. [doi: 10.1145/1134271.1134277]
22. Badawy A, Ferrara E, Lerman K. Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. In: Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2018 Presented at: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; August 28-31, 2018; Barcelona, Spain p. 258-265. [doi: 10.1109/asonam.2018.8508646]
23. Multidimensional-Ideological-Polarization. GitHub. 2020. URL: https://github.com/ashwinshreyas96/Multidimensional-Ideological-Polarization [accessed 2021-05-28]
24. Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In: Proceedings of the 43rd Hawaii International Conference on System Sciences. 2010 Jan 05 Presented at: 43rd Hawaii International Conference on System Sciences; January 5-8, 2010; Koloa, Kauai, HI p. 1-10. [doi: 10.1109/hicss.2010.412]
25. Metaxas P, Mustafaraj E, Wong K, Zeng L, O'Keefe M, Finn S. What do retweets indicate? Results from user survey and meta-review of research. In: Proceedings of the 9th International AAAI Conference on Web and Social Media. 2015 Presented at: 9th International AAAI Conference on Web and Social Media; May 26-29, 2015; Oxford, UK p. 658-661 URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/download/10555/10467
26. Raghavan UN, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks. Phys Rev E 2007 Sep 11;76(3):036106-1-036106-11 [FREE Full text] [doi: 10.1103/physreve.76.036106]
27. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. J Mach Learn Res 2003 Jan;3:993-1022 [FREE Full text]
28. Conover MD, Ratkiewicz J, Francisco M, Goncalves B, Flammini A, Menczer F. Political polarization on Twitter. In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media. 2011 Jul Presented at: 5th International AAAI Conference on Weblogs and Social Media; July 17–21, 2011; Barcelona, Spain p. 89-96 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2847/3275
29. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. 2017 Presented at: 15th Conference of the European Chapter of the Association for Computational Linguistics; April 3-7, 2017; Valencia, Spain p. 427-431 URL: https://www.aclweb.org/anthology/E17-2068.pdf
30. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proceedings of the 1st International Conference on Learning Representations. 2013 Presented at: 1st International Conference on Learning Representations; May 2-4, 2013; Scottsdale, AZ p. 1-12 URL: https://arxiv.org/pdf/1301.3781
31. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: 2014 Conference on Empirical Methods in Natural Language Processing; October 25-29, 2014; Doha, Qatar p. 1532-1543 URL: https://www.aclweb.org/anthology/D14-1162.pdf [doi: 10.3115/v1/d14-1162]
32. Gupta P, Pagliardini M, Jaggi M. Better word embeddings by disentangling contextual n-gram information. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Jun Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; June 2-7, 2019; Minneapolis, MN p. 933-939 URL: https://www.aclweb.org/anthology/N19-1098.pdf [doi: 10.18653/v1/n19-1098]
33. Smith L, Zhu L, Lerman K, Kozareva Z. The role of social media in the discussion of controversial topics. In: Proceedings of the 2013 International Conference on Social Computing. 2013 Presented at: 2013 International Conference on Social Computing; September 8-14, 2013; Washington, DC p. 236-243. [doi: 10.1109/socialcom.2013.41]

Abbreviations

DARPA: Defense Advanced Research Projects Agency
LDA: latent Dirichlet allocation
LPA: label propagation algorithm
PLD: pay-level domain

Edited by C Basch; submitted 22.12.20; peer-reviewed by S Dietze, C García; comments to author 08.02.21; revised version received 01.03.21; accepted 14.04.21; published 14.06.21.

Please cite as:
Rao A, Morstatter F, Hu M, Chen E, Burghardt K, Ferrara E, Lerman K
Political Partisanship and Antiscience Attitudes in Online Discussions About COVID-19: Twitter Content Analysis
J Med Internet Res 2021;23(6):e26692
URL: https://www.jmir.org/2021/6/e26692
doi:10.2196/26692
PMID:34014831

Original Paper

Professional Social Media Usage and Work Engagement Among Professionals in Finland Before and During the COVID-19 Pandemic: Four-Wave Follow-Up Study

Reetta Oksa[1*], MSc; Markus Kaakinen[2*], PhD; Nina Savela[1*], MSc; Jari J Hakanen[3*], PhD, Prof Dr; Atte Oksanen[1*], PhD, Prof Dr

[1]Faculty of Social Sciences, Tampere University, Tampere, Finland
[2]Institute of Criminology and Legal Policy, Faculty of Social Sciences, University of Helsinki, Helsinki, Finland
[3]Work Ability and Working Careers, Finnish Institute of Occupational Health, Helsinki, Finland
[*]all authors contributed equally

Corresponding Author:
Reetta Oksa, MSc
Faculty of Social Sciences
Tampere University
Kalevantie 5
Tampere, 33100
Finland
Phone: 358 504377619
Email: reetta.oksa@tuni.fi

Abstract

Background: The COVID-19 pandemic has changed work life profoundly and concerns regarding the mental well-being of employees' have arisen. Organizations have made rapid digital advancements and have started to use new collaborative tools such as social media platforms overnight.

Objective: Our study aimed to investigate how professional social media communication has affected work engagement before and during the COVID-19 pandemic and the role of perceived social support, task resources, and psychological distress as predictors and moderators of work engagement.

Methods: Nationally representative longitudinal survey data were collected in 2019-2020, and 965 respondents participated in all 4 surveys. Measures included work engagement, perceived social support and task resources, and psychological distress. The data were analyzed using a hybrid linear regression model.

Results: Work engagement remained stable and only decreased in autumn 2020. Within-person changes in social media communication at work, social support, task resources, and psychological distress were all associated with work engagement. The negative association between psychological distress and work engagement was stronger in autumn 2020 than before the COVID-19 outbreak.

Conclusions: The COVID-19 pandemic has exerted pressure on mental health at work. Fostering social support and task resources at work is important in maintaining work engagement. Social media communication could help maintain a supportive work environment.

(J Med Internet Res 2021;23(6):e29036) doi:10.2196/29036

KEYWORDS

COVID-19; engagement; mental health; moderator; predictor; psychological distress; social media; social support; support; task resources; usage; work engagement

Introduction

The rapid spread of the COVID-19 pandemic has affected our lives and work profoundly [1,2]. The COVID-19 pandemic has pressurized organizations to make a rapid digital leap to remote work and thus challenged and cultivated employees' well-being [3,4]. In Europe, 37% of the employees began working remotely in March and April 2020, with Finland having the largest proportion of remote workers (59%) [3]. In 2019, prior to the COVID-19 pandemic, only 23% of people in Finland worked remotely from home or other locations regularly, and 14% did so occasionally; therefore, the leap has been enormous [5].

In remote work conditions during the COVID-19 pandemic, the use of digital tools and social media platforms has increased at work for information and document creation, sharing, and exchange and for video meetings and discussions [6]. These tools are often used for both work and nonwork purposes among colleagues and have been found to enhance ways of working, foster innovation, allow for learning new skills, enhance performance, foster social relationships and social support, organizational identification, enable job satisfaction, and work engagement [6-11]. However, there is currently a lack of research on their role during the pandemic.

Work engagement, a key positive motivational state of well-being at work, is a comprehensive and enduring positive mental state that employees experience at their workplace and consists of three dimensions: vigor (ie, high energy levels, mental resilience, and persistence), dedication (ie, a sense of significance and pride), and absorption (ie, deep concentration on work and challenges detaching from work) [12,13]. Work engagement among employees in Finland was favorable before the COVID-19 crisis: 63% experienced vigor, 64% experienced dedication, and 56% experienced absorption in their work often or always [5].

According to job demands-resources model, work engagement is particularly driven by job resources, which are positive psychological, physical, social, and organizational characteristics of work, such as a good organizational climate and social support from colleagues and supervisors, which help employees accomplish work goals and foster learning and personal growth [13,14]. Social support defined as emotional, informational, and instrumental support, which describes not only the functional importance of relationships, but also the quality of those relationships and social belonging, can be a great reciprocal resource, for example, in coping stress and enhancing self-efficacy [15-18]. Engaged employees are more likely to be proactive and productive in their work [19]. Furthermore, autonomy, possibility to engage in meaningful work, and opportunities to leverage their strengths and experience at work are important factors influencing employee engagement [20-21].

Based on the Conservation of Resources (COR) theory, people tend to obtain and protect valuable resources, and loss of resources plays a significant role in the development of psychological stress [22]. Work engagement, as an energetic resource that employees may possess, should be a key priority in organizations, as it can increase life satisfaction and can prevent employees from psychological distress, depression, anxiety, sickness absenteeism, and burnout [23-26]. Furthermore, work engagement has been associated with healthy cardiac autonomic activity and a low likelihood of disability pensions [27,28]. Notably, high levels of work engagement have also been associated with increased short-term psychological distress and with decreased psychological stress over time [29]. At the societal level, work engagement predicts less unemployment in the general population [27].

The COVID-19 pandemic, along with its associated increase in digital and remote work, has potentially transformed ways of working for good [30]. Prior literature indicates that in the digital work environment, employees appreciate the opportunity to influence their work and enjoy the freedom and flexibility to complete their tasks; thus, they experience agency and higher self-esteem [31]. Resources such as support from managers received on social media can prevent work-related psychological distress [32,33]. Recent studies on the COVID-19 pandemic have reported that personal resilience and organizational and social support can sustain employee well-being and prevent anxiety [34]. Low supervisor support can, in turn, predict lower well-being, including stress, exhaustion, and burnout [35]. Furthermore, a study on adults in the United States reported that psychological distress increased from 3.9% in 2018 to 13.6% in 2020 during the COVID-19 pandemic [36]. Indeed, employees in the medical field have reported increased psychological distress and decreased well-being owing to heightened demands and workloads [34,37].

According to the COR theory [22] resource gains (such as supervisor support) in themselves have only a modest effect on well-being, but instead acquire saliency in the context of resource loss. Thus, prolongation of the COVID-19 pandemic can be considered a resource threat for employees. It can be argued that perceived social support and task resources have been particularly important in autumn 2020 as social distancing policies had been implemented since spring 2020 [1,2], and normal social interactions and working practices have been highly limited for a prolonged time. Among the basic psychological needs, particularly relatedness (lack of social contacts) and competence (eg, reduced possibilities to effectively bring about desired effects and outcomes) have been affected [38].

Social media communication at work has increased during the COVID-19 pandemic [6], and prior evidence has shown that work-related social media communication can enhance occupational resources such as social support and organizational identification and moreover work engagement [8]. However, previous studies have also indicated that psychological distress is associated with decreased work engagement before [39] and during the COVID-19 pandemic in spring 2020 [40]. Thus far, little is known about the longitudinal associations between professional social media communication and work engagement or how professional social media communication has affected work engagement and employees' mental well-being during the COVID-19 pandemic.

This longitudinal study analyzed changes in work engagement among employees in Finland before and during the COVID-19 pandemic. Our study investigated whether changes in social media communication at work, perceived social support, task resources, and psychological distress are related to changes in employees' work engagement, especially at a time of a prolonged pandemic. We proposed the following hypotheses: (1) increased social media communication predicts an increase in work engagement; (2) increased perceived social support and task resources at work predict an increase in work engagement; (3) increased psychological distress predicts decreased work engagement; and (4) the association between work engagement and (i) social media communication, (ii) social support, and (iii) psychological distress have been stronger during rather than before the COVID-19 pandemic.

Methods

Participants and Procedure

Data from a 4-timepoint longitudinal survey on social media usage at work in Finland from 2019 to 2020 were acquired to represent the working population in Finland. The first survey was conducted in March to April 2019 (timepoint 1 [T1]; n=1817). The participants were recontacted in September to October 2019 (timepoint 2 [T2]; n=1318), March to April 2020 (timepoint 3 [T3]; n=1081), and September to October 2020 (timepoint 4 [T4]; n=1152). The fourth survey was sent to all original respondents, whereas the third survey was sent only to those who had responded to the second survey.

The final sample in this study (n=965; 45.08% female; mean age 44.97 years, SD 11.36 years) included respondents who answered all 4 surveys, and the response rate was 53.11%. We found no major bias when conducting nonresponse analyses and when comparing the sample with official census figures of the working population in Finland [8]. The sample encompassed all major occupational fields and covered all prominent areas of Finland [6]. Analyses focused on employees of working age (18-66 years) and those respondents who remained employed (n=868). Only those respondents who finished the whole survey were included in the final data set. The survey study involved no ethical issues according to the assessment of the Academic Ethics Committee of Tampere region in Finland. The survey was conducted in Finnish, and participation was voluntary. The research group designed the survey and collected data in collaboration with Norstat, whose web-based research panel was used to recruit participants.

Measures

Work-Related and Nonwork-Related Social Media Communication

We measured the frequency of social media usage for work-related communication by asking the question, "How often do you use social media to keep in touch with your colleagues or work community regarding work-related matters (eg, sharing information or agreeing on timetables)?" We measured the frequency of social media usage for nonwork-related communication by asking the question, "How often do you use social media to keep in touch with your colleagues or work community regarding nonwork-related matters?" Possible answers were 0="I don't use it," 1="less than weekly," 2="weekly," 3="daily," and 4="many times a day." Both social media communications were measured at every time point; that is, every 6 months.

Work Engagement

Work engagement is most often measured using the Utrecht Work Engagement Scale (UWES) [41]. The 9-item version of this scale, UWES-9, is used most often owing to its construct validity [42]. Example questionnaire items include the following: "At my work, I feel that I am bursting with energy" and "I feel happy when I am working intensely." Responses are scored on a scale ranging from 0="never" to 6="always/every day." All 3 dimensions of the UWES were summed up to create a composite variable with a range of 0-54 and the Cronbach α coefficient was measured for all timepoints, ranging from .95 to .96. Work engagement was measured at every timepoint; that is, every 6 months.

Perceived Social Support

Perceived social support at work was measured using 4 questions on social support received from colleagues, supervisors, and the work community in general. These questions originate from the second version of the Copenhagen Psychosocial Questionnaire (Multimedia Appendix 1) [43], and they have been previously validated as a measure for social support at work [8]. Scores associated with these 4 items were summed to obtain a composite variable with a range of 4-20. Higher figures indicate higher perceived social support. The scale showed high reliability (Cronbach α=.74-.79). Perceived social support was measured at every timepoint; that is, every 6 months.

Task Resources

Task resources were measured using 4 questions from the work organization and job content dimension of the second version of the Copenhagen Psychosocial Questionnaire (Multimedia Appendix 2) [43]. Scores associated with the 4 questions were summed to obtain a composite variable with a range of 4-20. The scale showed adequately high internal consistency (Cronbach α=.67-.69). Task resources were measured at every time point; that is, every 6 months.

Psychological Distress

We measured psychological distress using the 12-item General Health Questionnaire [44]. Example questions included the following: "Have you recently felt constantly under strain?" and "Have you recently felt capable of making decisions about things?" Scores associated with all items were summed to obtain a composite variable with a range of 0-36. Higher scores indicate higher psychological distress. The scale showed high reliability (Cronbach α=.89-.92) between measurement points. Psychological distress was measured at every timepoint; that is, every 6 months.

Background Variables

Sociodemographic variables considered herein included age, gender, and education. All background variables were assessed at every timepoint; that is, every 6 months.

Statistical Analyses

As descriptive statistics, we expressed data as mean (SD) values for continuous study variables and frequencies and proportions for categorical variables (Tables 1 and 2). In addition, SD values between measurements were calculated for the within-person–level variables. We also assessed correlations among our study variables measured at different timepoints (Multimedia Appendix 3).

For all our hypotheses, we analyzed whether the within-person variation in social media communication, perceived social support, task resources, and psychological distress predicted changes in work engagement. We tested our hypotheses using a hybrid (or within-between) linear regression model [45]. This method decomposes the association between the dependent variables and time-variant independent variables into within-person and between-person effects. This is carried out by adding the individual means of dependent variables (between-person effects) and individual deviations from the person means (within-person effects) into the model simultaneously. Between-person effects are then estimated as associations between the individual means of the dependent and independent variables. Within-person effects are estimated as associations between the dependent variable and the observed deviation from the individual means. Thus, the between-person effects describe static differences between individuals, whereas within-person effects describe a dynamic relationship between the timely fluctuations in both the dependent and independent variables.

Table 1. Descriptive statistics of the study variables: continuous variables.

Variables		Time				Within-person differences, SD
	Range	T1, mean (SD)	T2, mean (SD)	T3, mean (SD)	T4, mean (SD)	
Work engagement	0-54	38.78 (12.13)	39.08 (12.15)	39.29 (11.64)	38.42 (12.04)	5.35
Work-related social media communication	0-4	1.27 (1.21)	1.31 (1.19)	1.52 (1.21)	1.51 (1.25)	0.69
Nonwork-related social media communication	0-4	1.16 (1.06)	1.10 (0.99)	1.24 (1.06)	1.18 (1.01)	0.59
Social support	4-20	14.65 (2.86)	14.56 (2.87)	14.68 (2.91)	14.65 (3.01)	1.49
Task resources	4-20	13.89 (2.76)	13.98 (2.74)	14.03 (2.63)	13.90 (2.70)	1.31
Psychological distress	12-48	24.89 (6.21)	24.14 (5.60)	24.26 (5.29)	24.19 (5.53)	3.32
Age in T1 (years)	18-64	43.52 (10.86)	N/A[a]	N/A	N/A	N/A

[a] N/A: not applicable.

Table 2. Descriptive statistics of the study variables: categorical variables.

Variables	Values	
	Coding	Number of participants, n (%)
Females[a]	0/1	379 (43.7)
Basic education	0/1	26 (3.0)
Secondary degree	0/1	429 (49.4)

[a] Number of participants at each time point (T1-T4)=868; total number of observations (T1+T2+T3+T4)=3472.

Our analysis proceeded in 2 steps. Model 1 included all our within-person and between-person main effects and a random intercept. For work-related and nonwork-related social media communication, perceived social support, task resources, and psychological distress, the effects were estimated as within- and between-person effects. For time, we estimated only within-person effects. Time was included as binary variables (T2-T4) with T1 as a reference category. Gender, age, and education at T1 were added to the model as between-person variables, as they varied only between persons.

To test our hypothesized moderation effects, within-person interaction terms including work-related and nonwork-related social media communication, perceived social support, task resources, and psychological distress at T4 were added to the model; Schunck [46] has described the estimation of within-person interaction terms. The significant interaction terms (95% CI) are reported in Model 2 in Tables 3 and 4. We report unstandardized regression coefficients (B), their estimated SE values, significance (*P* value), the variance of random intercept, and a log pseudolikelihood estimate in Tables 3 and 4. For effect size estimates, we reported Cohen f^2 coefficients for all the significant predictors. These coefficients were calculated using the approach described by Selya et al [47] and they can be interpreted as the proportion of explained variance associated with certain independent variables [48].

Table 3. Within-between models predicting changes in work engagement with time: fixed effects.

Variables	Model 1		Model 2	
	B (SE)	*P* value	B (SE)	*P* value
Constant	4.84 (3.82)	.21	3.99 (3.89)	.31
Within-person variables				
T2 (reference: T1)	0.08 (0.26)	.75	0.11 (0.26)	.68
T3 (reference: T1)	0.08 (0.28)	.77	0.10 (0.28)	.71
T4 (reference: T1)	*–0.66*[a]*(0.29)*	*.02*	2.69 (1.31)	*.04*
Work-related social media communication	*0.38 (0.15)*	*.009*	*0.38 (0.15)*	*.01*
Nonwork-related social media communication	0.11 (0.17)	.50	0.12 (0.17)	.48
Social support	*0.82 (0.09)*	*<.001*	*0.81 (0.09)*	*<.001*
Task resources	*0.91 (0.10)*	*<.001*	*0.92 (0.10)*	*<.001*
Psychological distress	*–0.28 (0.04)*	*<.001*	*–0.25 (0.04)*	*<.001*
Between-person variables				
Females	*4.02 (0.54)*	*<.001*	*4.02 (0.54)*	*<.001*
Basic education	–1.97 (1.87)	.29	–1.97 (1.87)	.29
Secondary degree	–0.11 (0.54)	.84	–0.11 (0.54)	.84
Age at T1 (years)	*0.08 (0.02)*	*.003*	*0.08 (0.02)*	*.003*
Work-related social media communication	0.44 (0.39)	.26	0.44 (0.39)	.26
Nonwork-related social media communication	*1.35 (0.45)*	*.003*	*1.35 (0.45)*	*.003*
Social support	*0.72 (0.14)*	*<.001*	*0.72 (0.14)*	*<.001*
Task resources	*1.89 (0.15)*	*<.001*	*1.89 (0.15)*	*<.001*
Psychological distress	*–0.57 (0.08)*	*<.001*	*–0.57 (0.08)*	*<.001*
Within-level interactions				
Psychological distress at T4	N/A[b]	N/A	*–0.14 (0.05)*	*.012*

[a] Values in italics are significant.

[b] N/A: not applicable.

Table 4. Within-between models predicting changes in work engagement with time: random effects.

Variables	Model 1	Model 2
Intercept, variance (95% CI)	52.49 (45.37-60.73)	52.52 (45.40-60.76)
Log pseudolikelihood	–11753.96	–11748.44

Results

The results of descriptive statistical analysis are shown in Tables 1 and 2. There were no significant changes in work engagement in T1-T3; however, in T4, work engagement decreased (B=–0.66; *P*=.02) (Table 3). The effect size of this change was low (Cohen f^2<.01). Among the other within-person variables, an increase in work-related social media communication (B=0.38; *P*=.009), social support (B=0.82; *P*<.001), and task resources (B=0.91; *P*<.001) were associated with increased work engagement. Increased psychological distress, in turn, was associated with reduced work engagement (B=–0.28; *P*<.001). The variance in work engagement was mainly explained by social support (Cohen f^2=.06), task resources (Cohen f^2=.05), and psychological distress (Cohen's f^2=.04), and the effect size for work-related social media communication was low (Cohen f^2<.01).

Between-person differences in nonwork-related social media communication (B=1.35; Cohen f^2<.01; *P*=.003), social support (B=0.72; Cohen f^2<.01; *P*<.001), and task resources (B=1.89; Cohen f^2=.01; *P*<.001) were positively associated with average work engagement, yet they only explained a marginal share of the variance in work engagement. Between-person differences in psychological distress, in turn, were negatively associated with work engagement (B=–0.57; *P*<.001). The effect size for this association was low (Cohen f^2<.01). In addition, female gender (B=4.02; *P*<.001) and age (B=0.08; *P*=.003) were associated with between-person differences in work engagement.

This implies that females reported higher work engagement on average than males, and older respondents also had higher work engagement on average. However, the effect size was low both for gender (Cohen $f^2<.01$) and age (Cohen $f^2<.01$).

Among our moderations (model 2), only the interaction effect between T4 and psychological distress was significantly related to work engagement (B=–0.14; *P*=.012). As expected, the negative association between within-person differences in work engagement and psychological distress was stronger in autumn 2020 (B=–0.39) than at T1 (B=–0.25; *P*<.001). However, the overall proportion of the variance in work engagement explained by this interaction was low (Cohen $f^2<.01$).

Discussion

Principal Findings

This study longitudinally investigated how social media communication at work predicts work engagement. Our theoretical and empirical model was based on the job demands-resources model and COR theory and considered the role of social support and task resources at work, along with psychological distress. Our results show that work engagement remained stable and only decreased in autumn 2020. Within-person changes in social media communication at work, social support, task resources, and psychological distress were associated with work engagement. Moreover, work engagement decreased during autumn 2020 when psychological distress had a stronger negative association with work engagement compared to that before the COVID-19 outbreak.

Our findings partly support hypothesis 1 and fully support hypothesis 2, thus demonstrating that more intensive work-related social media communication and higher perceived social support and task resources are associated with higher within-person work engagement. Nonwork-related communication with colleagues, perceived social support, and task resources were associated with between-person work engagement. However, within-person changes in nonwork-related social media communication did not predict changes in work engagement. Women and older people experienced higher work engagement, as reported previously for individuals in Finland and Europe [49,50].

Increased psychological distress was associated with reduced within-person work engagement, thus supporting hypothesis 3. Our results do not support hypotheses 4-i and 4-ii as the associations between work engagement and social media communication, perceived social support, and task resources did not change during the COVID-19 pandemic. The results partly support hypothesis 4-iii because the within-person association between psychological distress and work engagement was stronger during the COVID-19 pandemic (ie, autumn 2020).

Comparison With Prior Work

Our study is timely and the first one to offer longitudinal evidence regarding internal and external social media communication, both work-related and nonwork-related, in organizations and the related well-being implications, before and during the COVID-19 pandemic. Our findings revealed that work engagement remained considerably state at the onset of the COVID-19 pandemic during spring 2020. Hence, our results provide interesting insights and are in contrast with those of prior studies reporting that major disasters usually provoke stress and reduce resources [22,51]. However, prolonged uncertain situations have detrimental effects on well-being [52], which our results also confirm.

Increased psychological distress was associated with reduced work engagement in the within-person model, which is in line with prior reports on stress and social media use [53,54]. Individuals experienced higher psychological distress and lower work engagement during the autumn 2020, when COVID-19 was already well-known, and the crisis was underway. Therefore, our results contribute to the current literature on crises and the use of information and communication technologies [55,56], which indicate that a continued crisis has a negative influence on employee well-being and provides further knowledge, especially on professional social media use during the COVID-19 pandemic.

The significant role of various job resources in work engagement construction has been established in prior studies and in the context of social media [8,13,14]. Our findings strengthen the role of job resources in boosting work engagement during the pandemic by demonstrating that an increase in perceived social support and task resources fosters within-person and between-person work engagement.

Our findings have practical implications for organizations by demonstrating that work engagement decreased during autumn 2020, while psychological distress was stronger at that timepoint. Employees continued to work under uncertain conditions in autumn 2020 with no certain signs of future relief. Thus, providing mental health support for employees in such situations is crucial. The importance of supervisor support in alleviating employees' emotional exhaustion and feelings of uncertainty regarding COVID-19 has been previously reported [57], which our findings also emphasize. Furthermore, our results indicate that work-related social media communication is associated with enhanced work engagement, thus explaining within-person variation. Hence, communication with colleagues via social media can also serve as an important job resource that supports employees' resources and vigor, as well as their dedication to and absorption in their work.

Increased nonwork-related social media communication did not explain within-person variation in work engagement. We observed only between-person differences because those with high nonwork-related social media communication also had a higher level of work engagement on average. Employees who use social media actively for informal communication are also the ones who engage more in their work. This is because when engaged, employees invest energy into their work-related roles and therefore behave more proactively [58] and have higher contextual performance; that is, an individual's propensity to behave in ways that facilitate the social and psychological context of an organization [59]. Furthermore, the association between informal social media communication and work engagement might be more complex. For example, prior literature has reported that the association between informal social media communication and work engagement is mediated through other factors such as social support and organizational identification [8].

Moreover, increased social support and task resources were related to enhanced within-person and between-person work engagement. The results emphasize the importance of supporting employees in using their expertise, maintaining a sense of meaningfulness, providing possibilities to influence their work content and load, and offering and receiving social support.

Strengths and Limitations

We used a longitudinal, nationally representative sample, which enabled the analysis of timepoints before and during the COVID-19 crisis and the related effects on well-being, which can regard as one of the strengths of this study. The response rate was high, and our survey included a very limited number of missing observations. The study design with work-related and nonwork-related social media communication was novel, and a similar longitudinal study has not been performed before. The study was conducted with employed people in Finland and did not examine the COVID-19 crisis cross-nationally. Because this was an observational study, the associations reported herein should not be directly interpreted as causal relationships. Some effect sizes were low, but effect sizes for the main results remained significant even though our model was adjusted for a number of factors. This study was also limited to self-reported information.

Conclusions

Work engagement decreased during autumn 2020 at a time when psychological distress had a stronger negative association with work engagement. Social media communication at work, perceived social support, and task resources were also associated with higher work engagement. Overall, work engagement remained relatively stable during the COVID-19 crisis. However, providing mental health support during a prolonged crisis is crucial for organizations. Moreover, supporting employees' resources at work is important in maintaining employee work engagement, in which social media communication can be of help.

Acknowledgments

The Finnish Work Environment Fund (Professional Social Media Use and Work Engagement Among Young Adults Project, project number 118055 PI: Atte Oksanen). Data collection was also partly funded by the Faculty of Social Sciences at the Tampere University, Finland.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Copenhagen Psychosocial Questionnaire II Interpersonal relations and leadership dimension.
[DOCX File , 23 KB - jmir_v23i6e29036_app1.docx]

Multimedia Appendix 2
Copenhagen Psychosocial Questionnaire II Work organization and job contents dimension.
[DOCX File , 13 KB - jmir_v23i6e29036_app2.docx]

Multimedia Appendix 3
Correlation matrix.
[DOCX File , 100 KB - jmir_v23i6e29036_app3.docx]

References

1. Oksanen A, Kaakinen M, Latikka R, Savolainen I, Savela N, Koivula A. Regulation and Trust: 3-Month Follow-up Study on COVID-19 Mortality in 25 European Countries. JMIR Public Health Surveill 2020 Apr 24;6(2):e19218 [FREE Full text] [doi: 10.2196/19218] [Medline: 32301734]
2. Kniffin KM, Narayanan J, Anseel F, Antonakis J, Ashford SP, Bakker AB, et al. COVID-19 and the workplace: Implications, issues, and insights for future research and action. Am Psychol 2021 Jan;76(1):63-77. [doi: 10.1037/amp0000716] [Medline: 32772537]
3. Living, working and COVID-19: First findings – April 2020. Eurofound. 2020. URL: http://hdl.voced.edu.au/10707/538187 [accessed 2020-12-12]
4. Kestilä L, Härmä V, Rissanen P. Covid19-epidemian vaikutukset hyvinvointiin, palvelujärjestelmään ja kansantalouteen : Asiantuntija-arvio, syksy 2020. Julkari. 2020. URL: http://urn.fi/URN:ISBN:978-952-343-578-0 [accessed 2020-12-12]
5. Keyriläinen M. Working conditions barometer 2019. Ministry of Economic Affairs and Employment. 2020. URL: http://urn.fi/URN:ISBN:978-952-327-541-6 [accessed 2020-12-13]

6. Oksanen A, Oksa R, Savela N, Mantere E, Savolainen I, Kaakinen M. COVID-19 crisis and digital stressors at work: A longitudinal study on the Finnish working population. Computers in Human Behavior 2021 Sep;122:106853 [FREE Full text] [doi: 10.1016/j.chb.2021.106853]
7. Nisar T, Prabhakar G, Strakova L. Social media information benefits, knowledge management and smart organizations. J Bus Res 2019 Jan;94:264-272 [FREE Full text] [doi: 10.1016/j.jbusres.2018.05.005]
8. Oksa R, Kaakinen M, Savela N, Ellonen N, Oksanen A. Professional social media usage: Work engagement perspective. New Media & Society 2020 May 22:146144482092193 [FREE Full text] [doi: 10.1177/1461444820921938]
9. Waizenegger L, McKenna B, Cai W, Bendz T. An affordance perspective of team collaboration and enforced working from home during COVID-19. Eur J Inf Syst 2020 Aug 12;29(4):429-442 [FREE Full text] [doi: 10.1080/0960085X.2020.1800417]
10. Mäntymäki M, Riemer K. Enterprise social networking: A knowledge management perspective. Int J Inf Manage 2016 Dec;36(6):1042-1052 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2016.06.009]
11. Robertson BW, Kee KF. Social media at work: The roles of job satisfaction, employment status, and Facebook use with co-workers. Comput Hum Behav 2017 May;70:191-196 [FREE Full text] [doi: 10.1016/j.chb.2016.12.080]
12. Schaufeli W, Salanova M, Gonzales-Roma V, Bakker A. The measurement of engagement and burnout: A two sample confirmative factor analytic approach. J Happiness Stud 2002;3:71-92 [FREE Full text] [doi: 10.1023/A:1015630930326]
13. Schaufeli W, Bakker A. Job demands, job resources, and their relationship with burnout and engagement: a multi-sample study. J Organiz Behav 2004 Mar 30;25(3):293-315 [FREE Full text] [doi: 10.1002/job.248]
14. Hakanen J, Schaufeli W, Ahola K. The Job Demands-Resources model: A three-year cross-lagged study of burnout, depression, commitment, and work engagement. Work & Stress 2008 Jul;22(3):224-241 [FREE Full text] [doi: 10.1080/02678370802379432]
15. Bakker A. An Evidence-Based Model of Work Engagement. Curr Dir Psychol Sci 2011 Aug 08;20(4):265-269 [FREE Full text] [doi: 10.1177/0963721411414534]
16. Cohen S, Wills T. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]
17. Ross CE, Cohen S, Syme SL. Social Support and Health. Contemporary Sociology 1986 Mar;15(2):273. [doi: 10.2307/2071725]
18. Schwarzer R, Knoll N. Functional roles of social support within the stress and coping process: A theoretical and empirical overview. Int J Psychol 2007 Aug;42(4):243-252. [doi: 10.1080/00207590701396641]
19. Xanthopoulou D, Bakker A, Demerouti E, Schaufeli W. Work engagement and financial returns: A diary study on the role of job and personal resources. J Occup Organ Psychol 2011;82(1):183-200 [FREE Full text] [doi: 10.1348/096317908X285633]
20. Kahn W, Fellows S. Employee engagement and meaningful work. In: Dik BJ, Byrne ZS, Steger MF, editors. Purpose and meaning in the workplace. Washington, DC: American Psychological Association; 2013:105-126.
21. Van Wingerden J, Van der Stoep J. The motivational potential of meaningful work: Relationships with strengths use, work engagement, and performance. PLoS One 2018;13(6):e0197599 [FREE Full text] [doi: 10.1371/journal.pone.0197599] [Medline: 29897935]
22. Hobfoll S. The Influence of Culture, Community, and the Nested ‐ Self in the Stress Process: Advancing Conservation of Resources Theory. Appl Psychol 2001 Dec 21;50(3):337-421 [FREE Full text] [doi: 10.1111/1464-0597.00062]
23. Innstrand S, Langballe E, Falkum E. A longitudinal study of the relationship between work engagement and symptoms of anxiety and depression. Stress Health 2012 Feb;28(1):1-10. [doi: 10.1002/smi.1395] [Medline: 22259153]
24. Schaufeli W, Taris T, van Rhenen W. Workaholism, Burnout, and Work Engagement: Three of a Kind or Three Different Kinds of Employee Well-being? Appl Psychol 2008 Apr;57(2):173-203 [FREE Full text] [doi: 10.1111/j.1464-0597.2007.00285.x]
25. Schaufeli W, Bakker A, Van Rhenen W. How changes in job demands and resources predict burnout, work engagement, and sickness absenteeism. J Organiz Behav 2009 Oct;30(7):893-917 [FREE Full text] [doi: 10.1002/job.595]
26. Hakanen J, Schaufeli W. Do burnout and work engagement predict depressive symptoms and life satisfaction? A three-wave seven-year prospective study. J Affect Disord 2012 Dec 10;141(2-3):415-424. [doi: 10.1016/j.jad.2012.02.043] [Medline: 22445702]
27. Hakanen J, Rouvinen P, Ylhäinen I. The Impact of Work Engagement on Future Occupational Rankings, Wages, Unemployment, and Disability Pensions—A Register-Based Study of a Representative Sample of Finnish Employees. Sustainability 2021 Feb 03;13(4):1626 [FREE Full text] [doi: 10.3390/su13041626]
28. Seppälä P, Mauno S, Kinnunen M, Feldt T, Juuti T, Tolvanen A, et al. Is work engagement related to healthy cardiac autonomic activity? Evidence from a field study among Finnish women workers. J Posit Psychol 2012 Mar;7(2):95-106 [FREE Full text] [doi: 10.1080/17439760.2011.637342]
29. Shimazu A, Schaufeli W, Kubota K, Watanabe K, Kawakami N. Is too much work engagement detrimental? PLoS One 2018;13:e0208684 [FREE Full text] [doi: 10.1371/journal.pone.0208684] [Medline: 30586369]
30. Davison RM. The Transformative Potential of Disruptions: A Viewpoint. Int J Inf Manage 2020 Dec;55:102149 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102149] [Medline: 32836628]

31. Dittes S, Richter S, Richter A, Smolnik S. Toward the workplace of the future: How organizations can facilitate digital work. Business Horizons 2019 Sep;62(5):649-661 [FREE Full text] [doi: 10.1016/j.bushor.2019.05.004]
32. Charoensukmongkol P. Effects of support and job demands on social media use and work outcomes. Comput Hum Behav 2014 Jul;36:340-349 [FREE Full text] [doi: 10.1016/j.chb.2014.03.061]
33. Oshio T, Inoue A, Tsutsumi A. Associations among job demands and resources, work engagement, and psychological distress: fixed-effects model analysis in Japan. J Occup Health 2018 May 25;60(3):254-262 [FREE Full text] [doi: 10.1539/joh.2017-0293-OA] [Medline: 29563368]
34. Labrague L, De Los Santos JAA. COVID-19 anxiety among front-line nurses: Predictive role of organisational support, personal resilience and social support. J Nurs Manag 2020 Oct;28(7):1653-1661 [FREE Full text] [doi: 10.1111/jonm.13121] [Medline: 32770780]
35. Evanoff B, Strickland J, Dale A, Hayibor L, Page E, Duncan J, et al. Work-Related and Personal Factors Associated With Mental Well-Being During the COVID-19 Response: Survey of Health Care and Other Workers. J Med Internet Res 2020 Aug 25;22(8):e21366 [FREE Full text] [doi: 10.2196/21366] [Medline: 32763891]
36. McGinty E, Presskreischer R, Han H, Barry CL. Psychological Distress and Loneliness Reported by US Adults in 2018 and April 2020. JAMA 2020 Jul 07;324(1):93-94 [FREE Full text] [doi: 10.1001/jama.2020.9740] [Medline: 32492088]
37. Chen Q, Liang M, Li Y, Guo J, Fei D, Wang L, et al. Mental health care for medical staff in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e15-e16 [FREE Full text] [doi: 10.1016/S2215-0366(20)30078-X] [Medline: 32085839]
38. Ryan R, Deci E. Self-Determination Theory Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY: Guilford Publications; 2018.
39. Ramaci T, Pellerone M, Ledda C, Rapisarda V. Health promotion, psychological distress, and disease prevention in the workplace: a cross-sectional study of Italian adults. Risk Manag Healthc Policy 2017;10:167-175 [FREE Full text] [doi: 10.2147/RMHP.S139756] [Medline: 28860882]
40. Ruiz-Frutos C, Ortega-Moreno M, Allande-Cussó R, Ayuso-Murillo D, Domínguez-Salas S, Gómez-Salgado J. Sense of coherence, engagement, and work environment as precursors of psychological distress among non-health workers during the COVID-19 pandemic in Spain. Saf Sci 2021 Jan;133:105033 [FREE Full text] [doi: 10.1016/j.ssci.2020.105033] [Medline: 33046949]
41. Schaufeli W, Bakker A. UWES Utrecht work engagement scale. Preliminary manual Version 1. APA PsychNet 2006 [FREE Full text] [doi: 10.1037/t05561-000]
42. Seppälä P, Mauno S, Feldt T, Hakanen J, Kinnunen U, Tolvanen A, et al. The Construct Validity of the Utrecht Work Engagement Scale: Multisample and Longitudinal Evidence. J Happiness Stud 2008 May 6;10(4):459-481 [FREE Full text] [doi: 10.1007/s10902-008-9100-y]
43. Pejtersen J, Kristensen TS, Borg V, Bjorner JB. The second version of the Copenhagen Psychosocial Questionnaire. Scand J Public Health 2010 Feb;38(3 Suppl):8-24. [doi: 10.1177/1403494809349858] [Medline: 21172767]
44. Goldberg D, Gater R, Sartorius N, Ustun T, Piccinelli M, Gureje O, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997 Jan;27(1):191-197. [doi: 10.1017/s0033291796004242] [Medline: 9122299]
45. Schunck R, Perales F. Within- and Between-cluster Effects in Generalized Linear Mixed Models: A Discussion of Approaches and the Xthybrid command. The Stata Journal 2017 Mar 01;17(1):89-115 [FREE Full text] [doi: 10.1177/1536867X1701700106]
46. Schunck R. Within and between Estimates in Random-Effects Models: Advantages and Drawbacks of Correlated Random Effects and Hybrid Models. The Stata Journal 2013 Mar 01;13(1):65-76 [FREE Full text] [doi: 10.1177/1536867X1301300105]
47. Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ. A Practical Guide to Calculating Cohen's f(2), a Measure of Local Effect Size, from PROC MIXED. Front Psychol 2012;3:111 [FREE Full text] [doi: 10.3389/fpsyg.2012.00111] [Medline: 22529829]
48. Lorah J. Effect size measures for multilevel models: definition, interpretation, and TIMSS example. Large-scale Assess Educ 2018 Jul 23;6(1):8 [FREE Full text] [doi: 10.1186/s40536-018-0061-2]
49. Hakanen J, Ropponen A, Schaufeli W, De Witte H. Who is Engaged at Work?: A Large-Scale Study in 30 European Countries. J Occup Environ Med 2019 May;61(5):373-381. [doi: 10.1097/JOM.0000000000001528] [Medline: 30557226]
50. Hakanen J. Työn imun arviointimenetelmä. Työn imu-menetelmän (Utrecht Work Engage-ment Scale) käyttäminen, validointi ja viitetiedot Suomessa Utrecht Work Engagement Scale: Finnish Manual, Validation and Reference Data 2009.
51. Freedy J, Saladin M, Kilpatrick D, Resnick H, Saunders B. Understanding acute psychological distress following natural disaster. J Trauma Stress 1994 Apr;7(2):257-273. [doi: 10.1007/BF02102947] [Medline: 8012746]
52. Rettie H, Daniels J. Coping and tolerance of uncertainty: Predictors and mediators of mental health during the COVID-19 pandemic. Am Psychol 2020 Aug 03. [doi: 10.1037/amp0000710] [Medline: 32744841]
53. Maier C, Laumer S, Weinert C, Weitzel T. The effects of technostress and switching stress on discontinued use of social networking services: a study of Facebook use. Info Systems J 2015 Mar 25;25(3):275-308 [FREE Full text] [doi: 10.1111/isj.12068]

54. Salo M, Pirkkalainen H, Koskelainen T. Technostress and social networking services: Explaining users' concentration, sleep, identity, and social relation problems. Info Systems J 2018 Aug 06;29(2):408-435 [FREE Full text] [doi: 10.1111/isj.12213]
55. Leidner D, Pan G, Pan S. The role of IT in crisis response: Lessons from the SARS and Asian Tsunami disasters. J Strateg Inf Syst 2009 Jun;18(2):80-99 [FREE Full text] [doi: 10.1016/j.jsis.2009.05.001]
56. Leong C, Pan S, Ractham P, Kaewkitipong L. ICT-Enabled Community Empowerment in Crisis Response: Social Media in Thailand Flooding 2011. JAIS 2015 Mar;16(3):174-212 [FREE Full text] [doi: 10.17705/1jais.00390]
57. Charoensukmongkol P, Phungsoonthorn T. The effectiveness of supervisor support in lessening perceived uncertainties and emotional exhaustion of university employees during the COVID-19 crisis: the constraining role of organizational intransigence. J Gen Psychol 2020 Jul 21:1-20. [doi: 10.1080/00221309.2020.1795613] [Medline: 32691689]
58. Hakanen JJ, Peeters MCW, Schaufeli WB. Different types of employee well-being across time and their relationships with job crafting. J Occup Health Psychol 2018 Apr;23(2):289-301. [doi: 10.1037/ocp0000081] [Medline: 28191997]
59. Christian M, Garza A, Slaughter J. Work engagement: A quantitative review and test of its relations with task and contextual performance. Pers Psychol 2011;64(1):89-136 [FREE Full text] [doi: 10.1111/j.1744-6570.2010.01203.x]

Abbreviations

COR: Conservation of Resources
UWES: Utrecht Work Engagement Scale

Edited by C Basch; submitted 23.03.21; peer-reviewed by A Shimazu, M Attridge, C Escobar-Viera; comments to author 13.05.21; revised version received 28.05.21; accepted 28.05.21; published 15.06.21.

Please cite as:
Oksa R, Kaakinen M, Savela N, Hakanen JJ, Oksanen A
Professional Social Media Usage and Work Engagement Among Professionals in Finland Before and During the COVID-19 Pandemic: Four-Wave Follow-Up Study
J Med Internet Res 2021;23(6):e29036
URL: https://www.jmir.org/2021/6/e29036
doi:10.2196/29036
PMID:34048356

Original Paper

Social Media and Research Publication Activity During Early Stages of the COVID-19 Pandemic: Longitudinal Trend Analysis

Sonia L Taneja[1*], MD; Monica Passi[1*], MD; Sumona Bhattacharya[1], MD; Samuel A Schueler[1], MD; Sandeep Gurram[2], MD; Christopher Koh[3], MHSc, MD

[1]National Institutes of Diabetes and Digestive and Kidney Diseases, Digestive Disease Branch, Bethesda, MD, United States
[2]Urologic Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States
[3]Liver Diseases Branch, National Institutes of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, United States
[*]these authors contributed equally

Corresponding Author:
Christopher Koh, MHSc, MD
Liver Diseases Branch, National Institutes of Diabetes and Digestive and Kidney Diseases
10 Center Drive
Clinical Research Center, 5-2740
Bethesda, MD, 20892
United States
Phone: 1 301 443 9402
Email: christopher.koh@nih.gov

Abstract

Background: The COVID-19 pandemic has highlighted the importance of rapid dissemination of scientific and medical discoveries. Current platforms available for the distribution of scientific and clinical research data and information include preprint repositories and traditional peer-reviewed journals. In recent times, social media has emerged as a helpful platform to share scientific and medical discoveries.

Objective: This study aimed to comparatively analyze activity on social media (specifically, Twitter) and that related to publications in the form of preprint and peer-reviewed journal articles in the context of COVID-19 and gastroenterology during the early stages of the COVID-19 pandemic.

Methods: COVID-19–related data from Twitter (tweets and user data) and articles published in preprint servers (bioRxiv and medRxiv) as well as in the PubMed database were collected and analyzed during the first 6 months of the pandemic, from December 2019 through May 2020. Global and regional geographic and gastrointestinal organ–specific social media trends were compared to preprint and publication activity. Any relationship between Twitter activity and preprint articles published and that between Twitter activity and PubMed articles published overall, by organ system, and by geographic location were identified using Spearman's rank-order correlation.

Results: Over the 6-month period, 73,079 tweets from 44,609 users, 7164 journal publications, and 4702 preprint publications were retrieved. Twitter activity (ie, number of tweets) peaked in March 2020, whereas preprint and publication activity (ie, number of articles published) peaked in April 2020. Overall, strong correlations were identified between trends in Twitter activity and preprint and publication activity ($P<.001$ for both). COVID-19 data across the three platforms mainly concentrated on pulmonology or critical care, but when analyzing the field of gastroenterology specifically, most tweets pertained to pancreatology, most publications focused on hepatology, and most preprints covered hepatology and luminal gastroenterology. Furthermore, there were significant positive associations between trends in Twitter and publication activity for all gastroenterology topics (luminal gastroenterology: $P=.009$; hepatology and inflammatory bowel disease: $P=.006$; gastrointestinal endoscopy: $P=.007$), except pancreatology ($P=.20$), suggesting that Twitter activity did not correlate with publication activity for this topic. Finally, Twitter activity was the highest in the United States (7331 tweets), whereas PubMed activity was the highest in China (1768 publications).

Conclusions: The COVID-19 pandemic has highlighted the potential of social media as a vehicle for disseminating scientific information during a public health crisis. Sharing and spreading information on COVID-19 in a timely manner during the pandemic has been paramount; this was achieved at a much faster pace on social media, particularly on Twitter. Future investigation could demonstrate how social media can be used to augment and promote scholarly activity, especially as the world begins to increasingly rely on digital or virtual platforms. Scientists and clinicians should consider the use of social media in augmenting public awareness regarding their scholarly pursuits.

(J Med Internet Res 2021;23(6):e26956) doi:10.2196/26956

KEYWORDS

coronavirus; COVID-19; social media; gastroenterology; SARS-CoV-2; research; literature; dissemination; Twitter; preprint

Introduction

COVID-19, caused by the novel coronavirus SARS-CoV-2 (severe acute respiratory syndrome coronavirus), emerged into public view in December 2019 and resulted in a pandemic that has affected six continents, and it continues to indiscriminately affect individuals of all ages, races, and ethnicities. According to the World Health Organization, there have been more than 33,000,000 confirmed cases of COVID-19 globally, including more than 1,000,000 deaths reported by the end of September 2020—only 9 months after its emergence [1]. With time, different countries faced surges in cases straining their health care systems in unprecedented ways. It is during these times that the rapid dissemination of information related to this highly contagious virus and its management has been crucial.

Even though initial experiences related to COVID-19 primarily described respiratory complications, reports of gastrointestinal/gastroenterology (GI) involvement became more evident with increased clinical experience [2]. Although the extent of GI involvement with COVID-19 was uncertain based on early published experiences, it was postulated that this could be substantial due to the identification of the entry mechanism of SARS-CoV-2 that utilizes the angiotensin-2 (ACE2) receptor pathway, which is found throughout the GI tract, liver, and pancreas. Given the pathogen's similarities to the coronavirus known to cause severe acute respiratory syndrome (SARS) and Middle East Respiratory Syndrome (MERS), investigators suspected that prior experiences with these preceding viruses could provide insight into the current COVID-19 pandemic. Thus, GI luminal manifestations, the involvement of the liver and pancreas, and the management of unique GI patient populations were all considered areas of clinical and research interest [3-5]. Considering the rapid spread of COVID-19 and the consequent interruption of health care services across multiple fronts, the publishing of international experience with COVID-19 along with frequent updates in clinical guidance documents have assisted the GI community in managing this novel disease [6-9]. Furthermore, in an effort to mitigate the spread of infection, endoscopists have encountered significant changes to endoscopic practices by adopting new preprocedure regulations, use of enhanced personal protective equipment, and the rearrangement of endoscopy units to facilitate social distancing [10,11]. Additionally, with the implementation of national "lockdowns," the ability to share clinical experiences, analyze medical data, and disseminate management strategies for COVID-19 has become reliant on electronic media [12]. During these unprecedented times, the medical community has increasingly utilized social media (eg, Twitter, Facebook, TikTok) for communication and to facilitate interdisciplinary discussion [13,14].

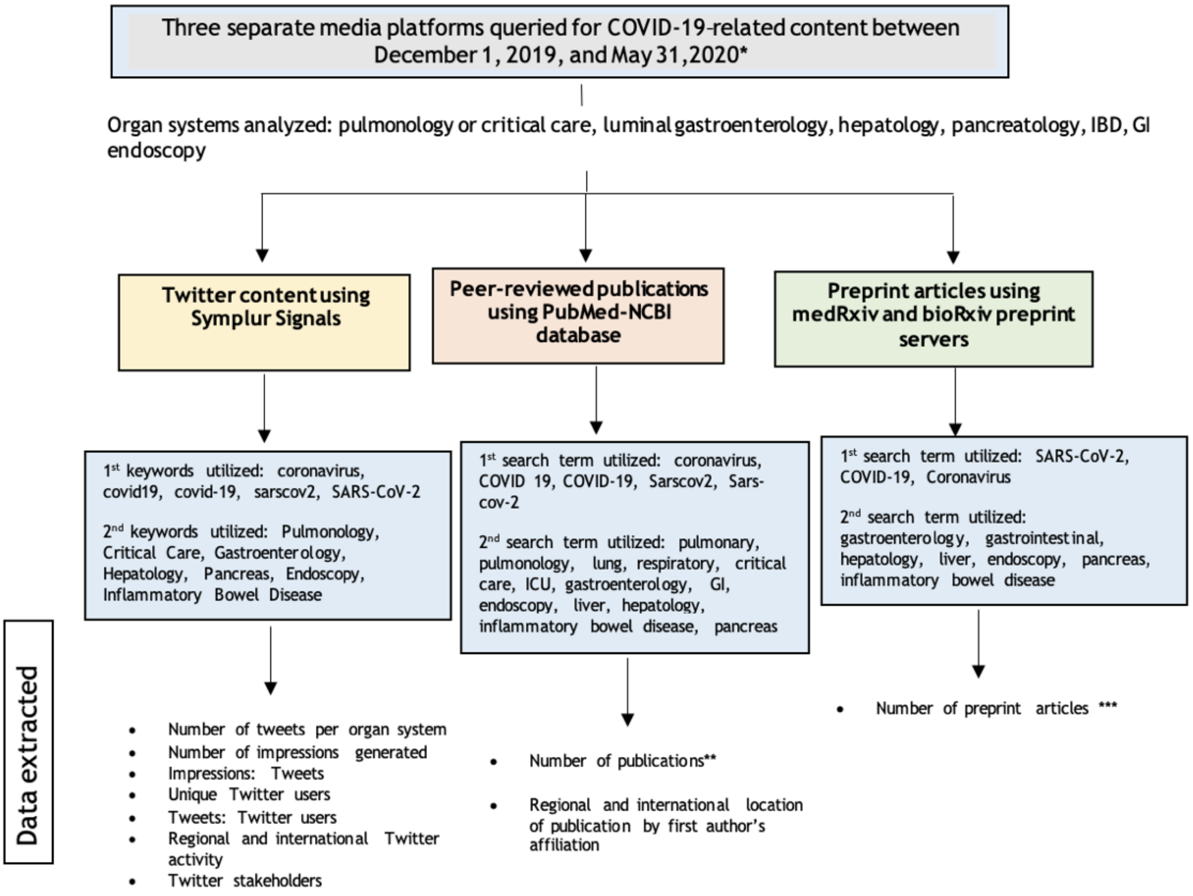
Social media platforms such as Twitter are social networking services through which any user or organization with an account, including those belonging to the scientific and medical communities, can share information and achievements. Compared to other professions, the health care community has been relatively reluctant in utilizing social media for professional purposes related to concerns on its potential impact on employment, medicolegal liability, and relationships among patients and colleagues [15-17]. Nonetheless, as these platforms continue to gain global acceptance and utilization, the ability to collect and analyze data from social media platforms has become essential in understanding health care–related needs, shifting public health interests, and highlighting areas for further medical study [18]. In this study, we aimed to explore COVID-19–related social media activity pertaining to the fields of gastroenterology and hepatology during the initial 6-month period of the COVID-19 pandemic, when knowledge about the virus was new and limited. Furthermore, we aimed to compare activity on social media—specifically Twitter—with that via more traditional channels of medical information sharing and distribution such as publications in medical journals and preprint repositories.

Methods

Data Collection for COVID-19–Related Twitter Activity

Data were collected using the publicly available Twitter analytics platform Symplur Signals [19]—a health care social media analytics platform that utilizes algorithms with natural language processing to provide in-depth information on Twitter activity. Data on topics related to COVID-19 were collected by performing specific searches categorizing the topics by organ system (Multimedia Appendix 1). Data were captured over the first 6 months of the COVID-19 pandemic, from December 1, 2019, to May 31, 2020. In an effort to capture the longitudinal evolution of social media use during this period, each month was split into half (ie, day 1-15 and day 16 to the end of the month). Data collected included total number of tweets and retweets, total number of impressions, total number of users, and user data, including the place of origin—by country (globally) and by state (within the United States). The ratios of tweets per Twitter user and impressions per tweet were also calculated (Figure 1). Definitions of these terms can be found in Multimedia Appendix 1.

Figure 1. Study flow diagram outlining the data extraction method from 3 media platforms analyzed (Twitter, PubMed-NCBI, and medRxiv or bioRxiv). *Each month of the specific study period was split into half-month intervals for the purposes of analysis. **Duplicate publications from separate searches were individually reviewed and reorganized into the most appropriate subject group in order to eliminate potential for publications to be accounted for more than once. ***A follow-up review of preprint articles pertaining to COVID-19 and gastroenterology that ultimately resulted in formal peer-reviewed journal publications was performed for July 2020. IBD: inflammatory bowel disease; GI: gastrointestinal.



Data Collection for COVID-19–Related Preprints and Publications

Preprint articles are research manuscripts shared publicly before peer review, which allows for rapid dissemination of information, thereby helping to inform policy and clinical practice in a timely manner. Preprint repositories have gained considerable attention over the course of the COVID-19 pandemic and have been increasingly utilized for the dissemination of crucial pandemic-related research. For our analysis, preprint articles related to COVID-19 biomedical research were identified using two popular preprint servers: medRxiv and bioRxiv. Specific search terms (see Multimedia Appendix 1) were used to identify and extract COVID-19 preprint articles for each half-month period across the 6-month study period for comparison with Twitter data. Furthermore, a follow-up review of the preprint articles pertaining to COVID-19 and gastroenterology that ultimately resulted in formal peer-reviewed journal publications was performed for the month of July 2020. This was done to account for the delay associated with the formal publication of a preprint article.

For the analysis of peer-reviewed publications, the PubMed–NCBI database was used to search for all publications pertaining to COVID-19 over the 6-month study period. The specific search terms used are detailed in Multimedia Appendix 1. All citations resulting from PubMed searches were recorded, and the search results were further filtered by half-month time intervals, identical to the search methods used for Twitter content and preprint articles for the purposes of comparison. For both preprints and publications, articles were further subgrouped by organ system topic. Duplicate publications from separate searches were individually reviewed and recategorized into the most appropriate subject group, thereby eliminating the potential for publications to be accounted for more than once. Finally, for each publication, the geographic location of the first author's institution was recorded.

Analysis of Social Media, Preprint, and Publication Activity

The primary outcome of the analysis was to identify the peak activity across the 3 platforms, as this shows how efficiently one platform could disseminate information as compared to the others. "Activity" was defined by the number of tweets (via Twitter), or publications (when referring to preprint or publication databases, such as bioRxiv and medRxiv or PubMed, respectively) produced. "Dissemination" could be defined in different ways as referenced by a previous systemic review report published by the Agency for Healthcare Research and Quality (AHRQ); however, for the purpose of this study, it was defined as “the active and targeted distribution of information or interventions via determined channels using planned strategies to a specific public health or clinical practice audience” [20]. In this study, the information distributed is the content of tweets or articles published in preprint repositories or accessible via the PubMed database. Of note, Twitter impressions is a convenient way to measure the exact distribution of the tweet content, as it calculates how many users would have been sent a particular tweet based on the number of followers the user who posted the tweet originally had. Secondary outcomes included (1) peak activity in each platform overall and then by GI subtopic, (2) peak activity in each platform by geographic location, and (3) comparison of trends between the different platforms overall, as well as by GI subtopic (see Multimedia Appendix 1).

Summary statistics of baseline data for tweets, impressions, preprints and PubMed publications are presented as frequencies for categorical data unless otherwise specified. Spearman’s rank-order correlation was performed to determine the relationship between Twitter activity (ie, tweets and impressions) and PubMed publications overall, by organ system and geographic location, as well as by Twitter activity and preprint articles overall and by organ system. Analysis was performed using STATA 15 (StataCorp, LLC) software. Statistical significance was set at *P*<.05. All authors had access to the study data and reviewed and approved the final manuscript.

Results

COVID-19–Related Publications and Twitter Activity Trends

Over the 6-month study period from December 1, 2019, through May 31, 2020, 73,079 tweets were identified from a total of 44,609 users, generating 207,039,610 impressions on the topic of COVID-19. During this same period, 7164 publications pertaining to COVID-19 were found to be indexed in PubMed along with 4702 preprints archived in medRxiv and bioRxiv repositories. The overall summary of Twitter and research publication activity by half-month time interval is shown in Table 1. Twitter activity, with regard to original tweets on the topic of COVID-19 did not appear until the latter half of January 2020, which resulted in 245 original tweets. This activity progressively increased thereafter and peaked during March 16-31, 2020, with 20,660 original tweets before gradually decreasing over the remaining study interval (Figure 2).

Table 1. A comparison of productivity trends by half-month intervals of COVID-19–related publication, preprint, and Twitter activity from December 2019 through May 2020.

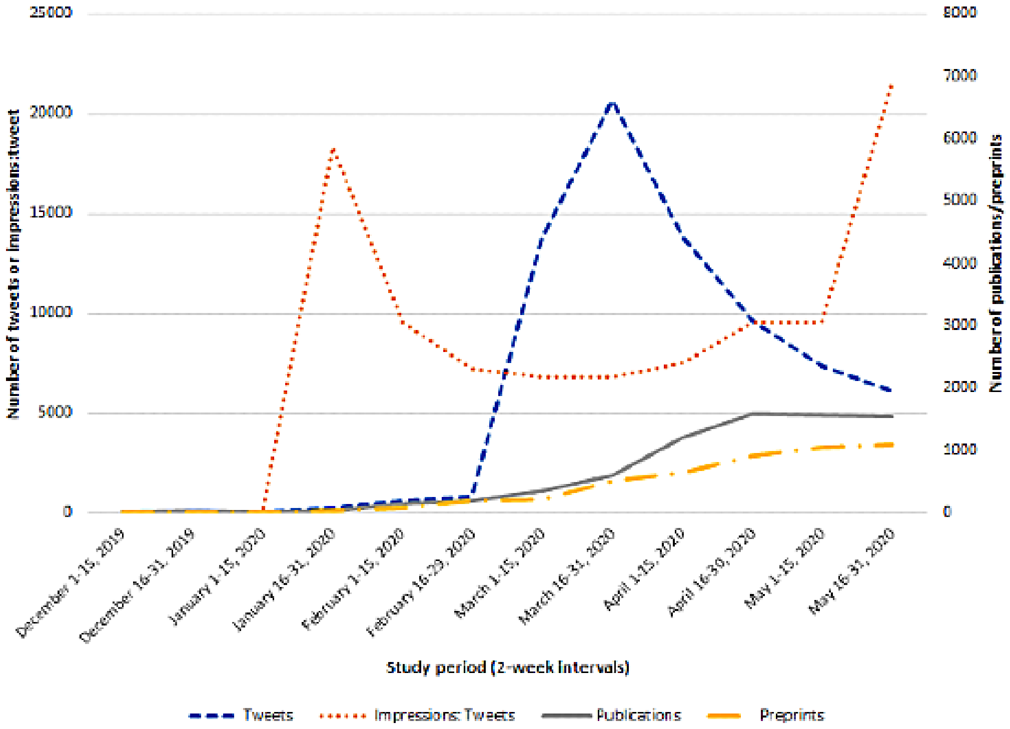
Informational sources	December 2019		January 2020		February 2020		March 2020		April 2020		May 2020	
	1-15	16-31	1-15	16-31	1-15	16-29	1-15	16-31	1-15	16-30	1-15	16-31
Publications[a], n	0	0	0	34	135	180	342	588	1196	1586	1561	1541
Preprints[b], n	0	0	0	35	71	181	217	506	633	922	1051	1086
Tweets[c], n	0	0	0	245	592	815	13797	20660	13845	9636	7399	6090
Impressions	0	0	0	1,439,197	1,809,224	180,594	30,061,305	44,640,303	33,351,337	29,411,077	22,536,326	41,900,257
Impressions per tweet	0	0	0	5874	3056	2320	2179	2179	2409	3052	3046	6880
Unique users	0	0	0	165	362	552	9727	13034	8210	5647	3759	3153
Tweets per user	0	0	0	1.48	1.64	1.48	1.42	1.59	1.69	1.71	1.97	1.93

[a] Publications indexed in PubMed–NCBI database.

[b] Preprints located in the medRxiv and bioRxiv repositories.

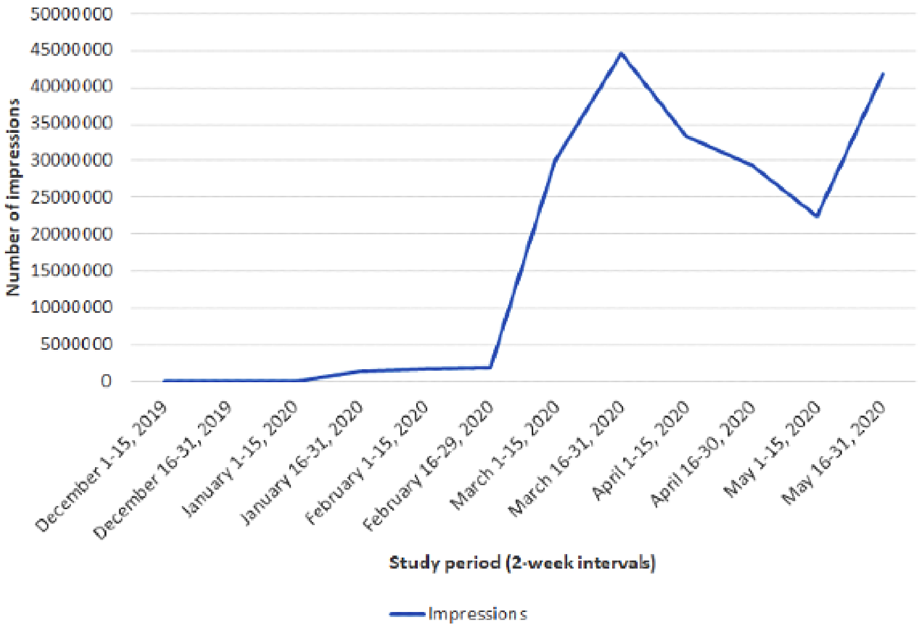
[c] Tweets and associated variables extracted using Symplur Signals search engine.

Figure 2. Trend of COVID-19–related tweets, ratio of impressions to tweets, publications and preprints. Twitter activity from December 1, 2019, through May 31, 2020, captured at half-month intervals by using the Twitter analytics platform Symplur Signals. Preprints from medRxiv and bioRxiv repositories were also abstracted during this time period along with publications indexed in the PubMed-NCBI database.



A similar pattern of activity was observed among the number of Twitter users who posted tweets related to COVID-19, which increased from 165 users to 13,034 users between January and the latter half of March 2020. Twitter impressions followed a similar pattern, with a peak observed during the second half of March 2020. Interestingly, a second peak in impressions was apparent during the latter half of May 2020, which was not observed with regard to the number of tweets and Twitter users (Figure 3). On average, the number of tweets per Twitter user ranged from 1.48 to 1.97. Impressions generated per tweet were initially high (5874 impressions/tweet) in the latter half of January 2020 but did not peak until the latter half of May 2020 (6880 impressions/tweet). Temporal trends are further detailed in Figure 2.

Figure 3. Trend of COVID-19 Twitter impressions. Number of impressions generated from Twitter from December 1, 2019, through May 31, 2020, captured at half-month intervals by using the Twitter analytics platform Symplur Signals.



Scientific COVID-19–related articles indexed in PubMed as well as preprints in medRxiv and bioRxiv servers followed a similar trajectory as Twitter activity, with both the first peer-reviewed articles and the first preprints becoming available during the second half of January 2020 [21]. However, unlike Twitter activity that peaked in the second half of March 2020, publications and preprints reached peak activity around the second half of April 2020. Notably, we observed a parallel rise in the number of preprints and PubMed publications (Table 1 and Figure 2).

A moderately strong correlation was demonstrated between Twitter activity (ie, number of tweets) and number of PubMed publications (ρ=0.58), as well as between Twitter activity and number of preprints (ρ=0.57) across the study duration (*P*<.001 for both; Table 2).

Similarly, there was a moderately strong association between the number of Twitter impressions and PubMed publications (ρ=0.56, *P*<.001), as well as between the number of Twitter impressions and preprints (ρ=0.54, *P*<.001; Table 3).

Table 2. Correlation between Twitter activity (tweets) and PubMed publications and between Twitter activity (tweets) and preprints by organ system. Italicized values indicate statistical significance.

Organ system	Tweets and PubMed publications (ρ)	*P* value	Tweets and preprints (ρ)	*P* value
Overall trend	0.58	*<.001*	0.57	*<.001*
Trend by organ system				
Pulmonology or critical care	0.8	*.002*	0.8	*.003*
Luminal gastroenterology	0.7	*.009*	0.6	*.03*
Hepatology	0.7	*.006*	0.7	*.009*
Inflammatory bowel disease	0.7	*.006*	0.5	*.07*
Pancreatology	0.4	.20	0.4	.30
Gastrointestinal endoscopy	0.7	*.007*	0.7	*.02*

Table 3. Correlation between Twitter activity (impressions) and PubMed publications and between Twitter activity (impressions) and preprints by organ system. Italicized values indicate statistical significance.

Organ system	Impressions and PubMed publications (ρ)	*P* value	Impressions and preprints (ρ)	*P* value
Overall trend	0.56	*<.001*	0.54	*<.001*
Trend by organ system				
Pulmonology or critical care	0.8	*.001*	0.8	*.002*
Luminal gastroenterology	0.7	*.009*	0.7	*.006*
Hepatology	0.7	*.005*	0.7	*.006*
Inflammatory bowel disease	0.8	*.004*	0.4	.20
Pancreatology	0.5	*.07*	0.3	.30
Gastrointestinal endoscopy	0.8	*.004*	0.7	*.02*

COVID-19–Related Twitter, Publication, and Preprint Content Classified by Organ System Topic

Overview

Analysis of Twitter, publication, and preprint data pertaining to the effects of COVID-19 on specific organ system topics are outlined in Table S1 of Multimedia Appendix 2, and trends are illustrated in Figure 4 and Figure 5. The majority of COVID-19–related tweets (58,792/73,079, 80.4%), publications (6713/7164, 93.7%), and preprint articles (4567/4702, 97.1%) covered the topic of pulmonology or critical care. Gastroenterology was a small subset of these topics; however, within the field, the majority of tweets were on the topics of pancreatology (5804/73,079, 7.9%), followed by luminal gastroenterology (3318/73,079, 4.5%), inflammatory bowel disease (2818/73,079, 3.9%), GI endoscopy (1764/73,079, 2.4%), and hepatology (583/73,079, 0.8%). With regard to gastroenterology-related publications, the majority of articles were on the topic of hepatology (236/7250, 3.3%) followed by GI endoscopy (111/7250, 1.5%), luminal gastroenterology (64/7250, 0.9%), IBD (30/7250, 0.4%), and pancreatology (10/7250, 0.1%). Preprint publications were primarily on the topics of luminal gastroenterology (21/4702, 0.4%), hepatology (18/4702, 0.4%), and IBD (3/4702, 0.1%) (see Table S1 in Multimedia Appendix 2).

Figure 4. Trend of Twitter, publication, and preprint activity related to COVID-19 and pulmonary or critical care topics. Comparison of the number of tweets posted with articles published in peer-reviewed journals and preprints published pertaining to COVID 19 and pulmonary or critical care at half-month intervals between December 2019 and May 2020.

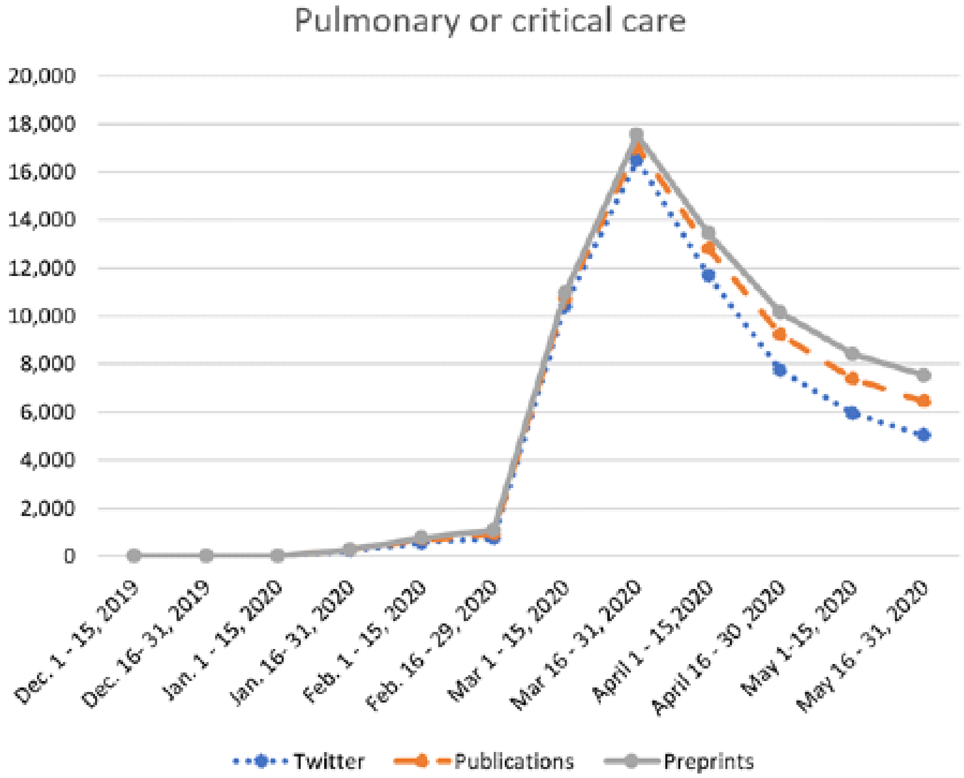
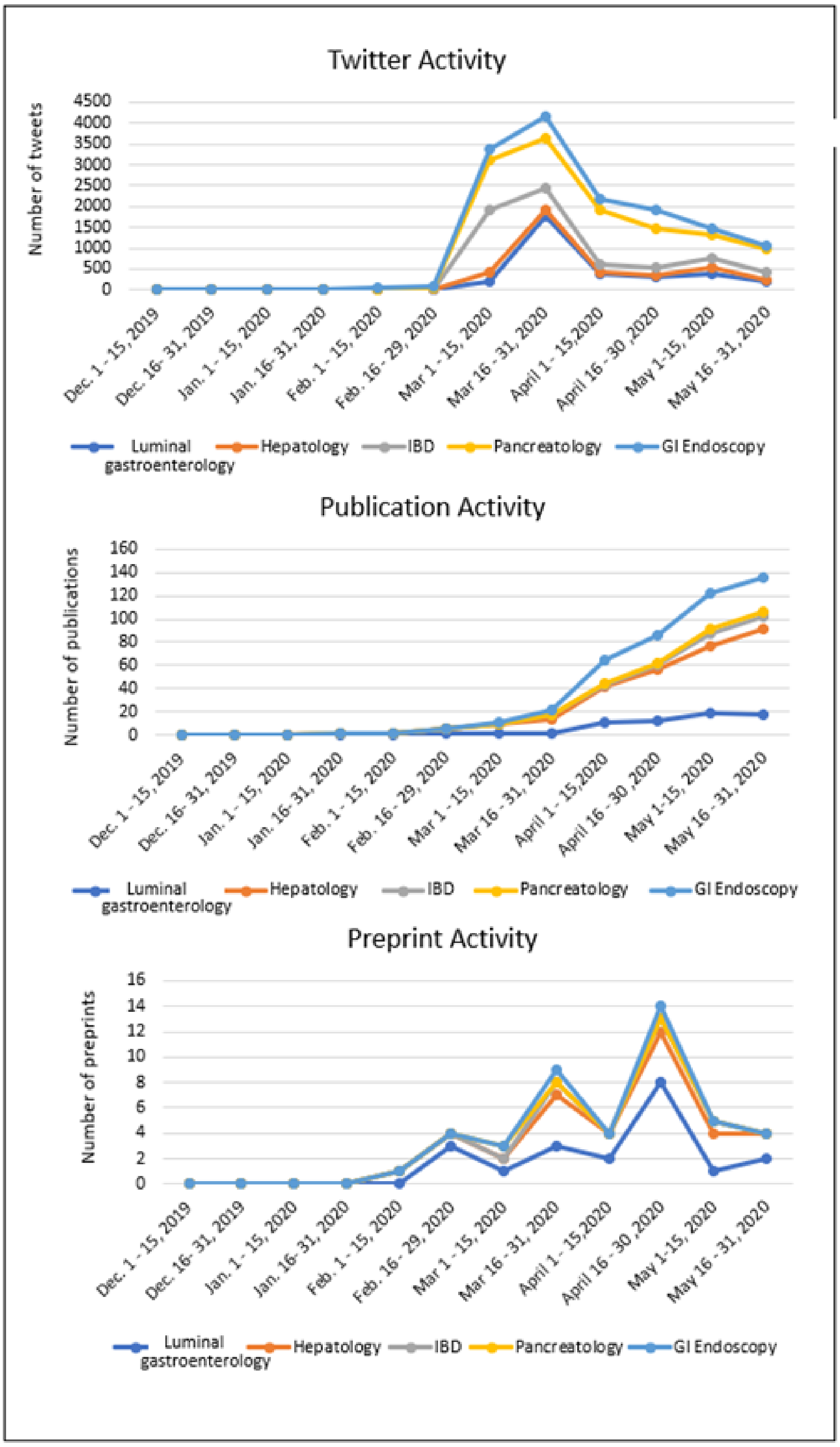


Figure 5. Trend of COVID-19 and gastroenterology subspecialty–related Twitter, publication, and preprint activity. Trend in the number of (A) tweets, (B) publications, and (C) preprints published pertaining to COVID 19 and gastroenterology subspecialty topics at half-month intervals between December 2019 and May 2020. GI: gastrointestinal; IBD: inflammatory bowel disease.



Pulmonology and Critical Care

Approximately 59,000 COVID-19–related tweets analyzed across the study period were on the topic of pulmonology or critical care. The most significant increase in tweets on this topic occurred between February and March 2020, with a nearly 22-fold increase, and an ultimate peak in activity (n=16,489) was observed in the latter half of March (Figure 4 and Table S1 in Multimedia Appendix 2).

A total of 6713 peer-reviewed articles on the topic of pulmonology or critical care and COVID-19 were indexed in PubMed during the 6-month study period, and these first appeared in January 2020. The most significant increase (ie, 2-fold) in the number of publications on COVID-19 and pulmonology or critical care was observed between the latter half of March 2020 and the start of April 2020. As compared to peer-reviewed publications, there were approximately one-third fewer preprint articles (n=4567) in medRxiv and bioRxiv related to COVID-19 and pulmonology or critical care identified during the study period. Moreover, on the topic of pulmonology or critical care and COVID-19, the longitudinal trend in preprint article availability appears to have paralleled publications indexed in PubMed; however, for preprint articles, the most significant rise was observed 2 weeks prior to that observed with PubMed publications, specifically between the first and second half of March (Figure 4 and Table S1 in Multimedia Appendix 2).

There was a strong correlation between both the number of COVID-19–related tweets and peer-reviewed publications (ρ=0.8, *P*=.002) as well as between the number of COVID-19–related tweets and preprints (ρ=0.8, *P*=.003, respectively; Table 2) on the topic of pulmonology or critical care. Similarly, there was a strong correlation between both pulmonology or critical care–related Twitter impressions and publications (ρ=0.8, *P*=.001) as well as between Twitter impressions and preprints (ρ=0.8, *P*=.002; Table 3).

Gastroenterology

A total of 14,285 tweets concerning the field of gastroenterology and COVID-19 (encompassing subspecialty fields of luminal gastroenterology, IBD, hepatology, GI endoscopy, and pancreatology) were identified during the study period (Figure 5 and Table S1 in Multimedia Appendix 2). Among all tweets recorded during the 6-month study period, 19.6% (14,287/73,079) were on the topic of COVID-19 and gastroenterology. The longitudinal trend in number of gastroenterology-related tweets (including subspecialty gastroenterology fields) paralleled that observed with pulmonology or critical care–related tweets, with an approximate 45-fold increase in the number of tweets spanning the latter half of February 2020 (n=75) and peaking in the second half of March 2020 (n=4171). When further stratified by subspecialty field, the majority of COVID-19 and gastroenterology–related tweets were on the topic of pancreatology (5804/14,287, 40.6%), followed by luminal gastroenterology (3318/14,287, 23.2%), IBD (2647/14,287, 19.7%), GI endoscopy (1764/14,287, 12.3%), and hepatology (583/14,287, 4.1%).

A total of 449 peer-reviewed publications related to COVID-19 and gastroenterology were identified in PubMed during the study period. In contrast to Twitter activity, the majority of these publications were on the topic of hepatology (235/449, 52.3%) followed by GI endoscopy (111/449, 24.6%), luminal gastroenterology (64/449, 14.2%), IBD (30/449, 6.7%), and pancreatology (10/449, 2.2%). Similar to Twitter activity, PubMed publications on the topics of luminal gastroenterology and hepatology first appeared in the latter half of January 2020. The most significant increase in COVID-19 and liver–related publications was observed between the latter half of March (n=12) and early April 2020 (n=31), with an over 2.5-fold increase in the number of publications on this topic. Luminal gastroenterology–related publications, which first appeared in the latter half of February 2020, significantly increased between the second half of March and early April 2020, with a 5-fold increase as detailed in Table S1 in Multimedia Appendix 2.

A total of 45 COVID-19 and gastroenterology–related preprints were archived in medRxiv and bioRxiv servers over the study period. Longitudinal analysis showed that the number of preprints on the topic of gastroenterology peaked in the latter half of April 2020. When further stratified by subspecialty, unlike that observed with peer-reviewed publications, the majority of preprints covered luminal gastroenterology (21/45, 46.7%), followed by hepatology (18/45, 40%), IBD (3/45, 6.7%), GI endoscopy (2/45, 4.4%), and pancreatology (1/45, 2.2%) (Table S1 in Multimedia Appendix 2).

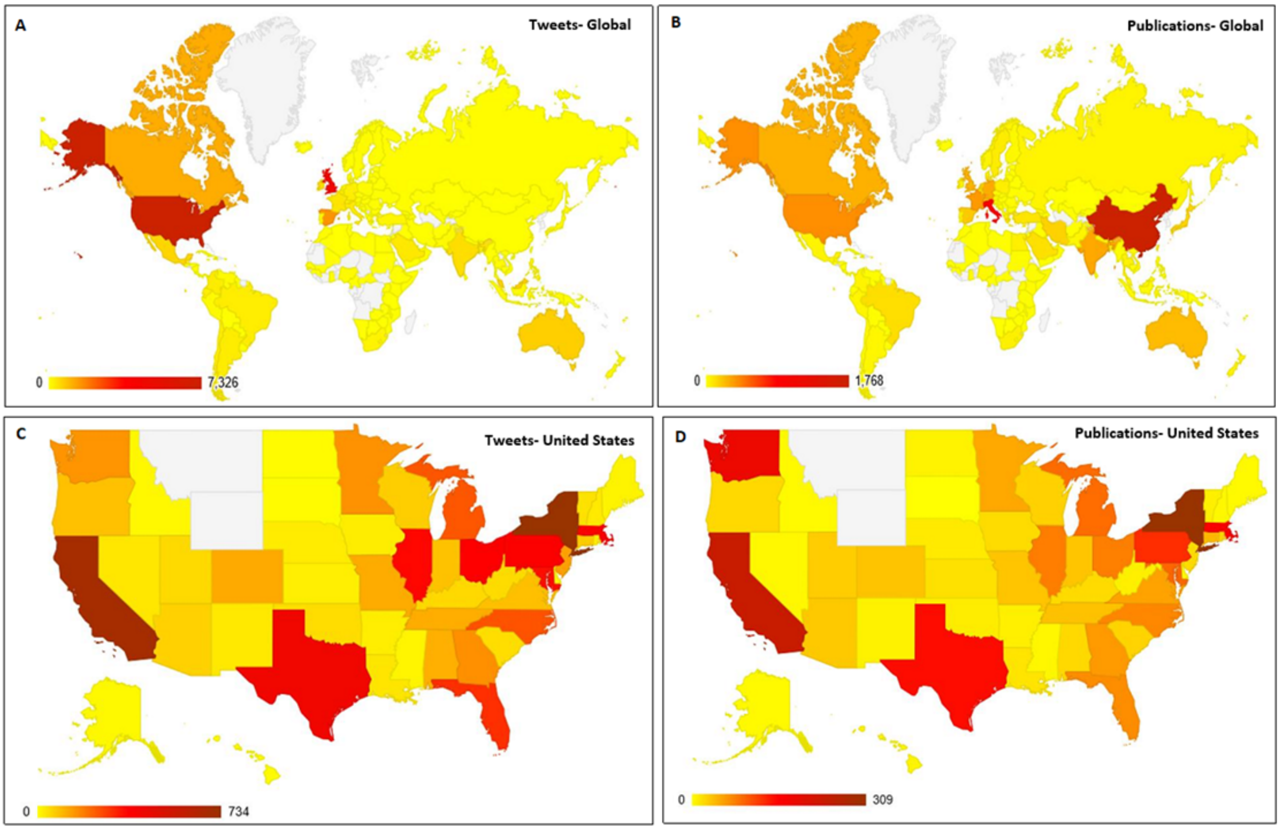
Similar to pulmonology or critical care–related content, there was a strong correlation between tweets and peer-reviewed publications (ρ=0.6, *P*=.03) as well as between tweets and preprints (ρ=0.7, *P*=.009) on the topic of luminal gastroenterology. Additionally, a strong correlation was identified between both the number of tweets and peer-reviewed publications (ρ=0.7, *P*=.006) as well as between tweets and preprints (ρ=0.7, *P*=.009) on the topic of COVID-19 and hepatology. A similarly strong correlation was observed between the number of tweets and PubMed publications on the topic of GI endoscopy (ρ=0.7, *P*=.007), the number of tweets and preprints on the topic of GI endoscopy (ρ=0.7, *P*=.02), and the number of tweets and peer-reviewed publications on the topic of COVID-19 and IBD (ρ=0.7, *P*=.008). In contrast, no significant correlation was identified between tweets and peer-reviewed publications or preprints on the topic of COVID-19 and pancreatology (Table 2).

Regarding COVID-19 and luminal gastroenterology content, a strong correlation was observed between both Twitter impressions and peer-reviewed publications (ρ=0.7, *P*=.009) as well as between Twitter impressions and preprints (ρ=0.7, *P*=.006). Similarly, strong correlations were identified between Twitter impressions and peer-reviewed publications on the topics of COVID-19 and hepatology (ρ=0.7, *P*=.005), IBD (ρ=0.8, *P*=.004), and GI endoscopy (ρ=0.8, *P*=.004). There was no significant correlation between Twitter impressions and publications on the topic of COVID-19 and pancreatology. In evaluating the association between Twitter impressions and preprints, strong associations were found on the topics of hepatology (ρ=0.7, *P*=.006) and GI endoscopy (ρ=0.7, *P*=.02), whereas no associations were found pertaining to the topics of IBD and pancreatology (Table 3).

COVID-19 Twitter and Publication Content by Geographic Location

The top 5 countries with the highest number of COVID-19–related tweets posted globally over the 6-month study period included the United States (7331/22,215, 33.0%), followed by the United Kingdom (4229/22,215, 19.3%), Spain (1527/22,215, 6.8%), Canada (1174/22,215, 5.3%), and Australia (673/22,215, 3.0%). China generated the highest number of peer-reviewed publications indexed in PubMed (1768/6352, 27.8%) throughout the study period, followed by Italy (915/6352, 14.4%), the United States (389/6352, 6.1%), France (348/6352, 5.5%), and India (303/6352, 4.8%). Figure 6 illustrates the countries with the highest Twitter activity and peer-reviewed publication activity. The top 20 countries with the highest number of tweets and the highest number of peer-reviewed publications are listed in Tables S3 and S4, respectively, in Multimedia Appendix 2. There was a strong correlation between the number of tweets and peer-reviewed publications in both the United States (ρ=0.8, *P*=.005) and the United Kingdom (ρ=0.8, *P*=.01; see Table S2 in in Multimedia Appendix 2).

Figure 6. Heat maps illustrating the total number of COVID-19–related tweets (A) across the globe and (C) in the United States, as well as publications indexed in the PubMed-NCBI represented (B) across the globe and (D) in the United States, over the 6-month study period (December 2019 through May 2020). Numbers are represented on the spectrum from the least (yellow) to the highest amount (maroon), as detailed in the legend accompanying each map. Countries or US states shaded white indicate the absence of data for those regions.



Within the United States, when analyzing both Twitter activity and peer-reviewed publications, the state of New York had the highest COVID-19–related activity (11% of tweets and 39.1% of publications) during the study period, followed closely by California (10.3% of tweets and 36.6% of publications). Figure 6 illustrates the Twitter activity and publication activity in the US states. The top 20 US states with the highest number of tweets and peer-reviewed publications are listed in Tables S5 and S6, respectively, in Multimedia Appendix 2.

COVID-19 Twitter Content by User Stakeholder Designation or Category

Twitter user data entered as the Twitter user's self-designated health care stakeholder role was analyzed. For the topics of pulmonology or critical care, luminal gastroenterology, hepatology, IBD, and GI endoscopy, the top 2 most active stakeholder categories were doctors/physicians and researchers/academic users. Furthermore, for the topic of pancreatology, advocacy organizations and patient advocates were the most active stakeholder users. The top 15 most active users and their stakeholder roles categorized by Twitter activity for each organ system topic are further detailed in Table S7 in Multimedia Appendix 2.

Discussion

Principal Findings

Since its emergence in December 2019, the novel coronavirus SARS-CoV-2 has triggered an unparalleled global response in the fields of science, medicine, public health, and technology. Considering its highly contagious nature, along with the paucity of knowledge and current lack of effective treatment modalities to combat the infection, the need for rapid sharing and dissemination of information has been paramount. In this study, we assessed the dissemination of COVID-19–related information via preprint services, formal peer-reviewed publications, and through the global reach of the social media platform Twitter.

Specifically, we observed that during the second half of March 2020, when COVID-19 was continuing to spread rapidly prompting various nations, including the United States, to enter a state of lockdown, social media activity on Twitter was at its peak, with almost 7000 impressions per tweet analyzed. Furthermore, we observed that Twitter activity was strongly correlated with the published scientific data available to the general public. Although COVID-19 has been predominantly linked with severe pulmonary complications, approximately 20% of the conversations on social media was related to the field of gastroenterology with specific discussions related to hepatology, GI endoscopy, luminal gastroenterology, IBD, and pancreatology. Social media activity was strongly associated with the availability of published data pertaining to all gastroenterology topics with the exception of pancreatology. Finally, in our analysis of data by geographic region of publication, social media activity was most prominent in the regions most affected by the pandemic both globally (in regions where social media via Twitter is not banned) and within the United States, with strong associations between social media and publication data.

In the longitudinal assessment of publication activity, the authors observed that peak social media activity predated peak PubMed publications by approximately 30 days. One possible explanation for this lag interval is the technical delay that typically occurs after acceptance of a peer-reviewed manuscript by a journal and prior to indexing in PubMed. It is worth highlighting that around the peak of PubMed publication activity, we also witnessed a parallel rise in the preprint repository activity. Historically, the use of preprint repositories has allowed researchers to “claim the space” or even to “publish first” in contentious fields of science and research. However, in the face of an evolving pandemic, preprint repositories are serving as a new mode of scientific communication, bypassing the typical lengthy peer-review process and thus allowing for faster dissemination and communication of research and clinical findings related to COVID-19. In fact, in March 2020, when the World Health Organization officially declared COVID-19 a pandemic, 8830 biomedical preprints were published, a 142% increase from the previous year, and medRxiv page views had increased to 15 million a month, as compared to 1 million a month prior to the start of the pandemic [22]. As of July 1, 2020, we were able to document that nearly half of all preprint articles related to COVID-19 and gastroenterology have subsequently made their way into reputable scientific journals, further supporting this theory.

Notably, we observed that physicians, nonmedical doctors, and scientific researchers constituted the lead stakeholder activity for social media use overall during the COVID-19 pandemic. This observation suggests that even during the early stages of the COVID-19 pandemic, social media became an increasingly sought-after tool, likely for the purpose of communicating medical and scientific information. This study confirms—by documenting health care stakeholder activity—that the scientific and medical communities leveraged social media platforms during the early stages of the COVID-19 pandemic more so than other health care stakeholders such as patients or advocacy groups. Previous studies performed during nonpandemic times have demonstrated conflicting evidence regarding the impact of social media on scientific publications related to citation impact, metrics, and viewability [23-26]. This study clearly illustrated that social media activity was greater and also peaked earlier than any of the publication modalities, which is important to consider when investigating how social media could potentially impact research and scientific work in future.

The gastroenterology subtopic of pancreatology is a good example of this potential influence, or rather, the lack thereof. Although pancreatic manifestations (eg, acute pancreatitis) of COVID-19 have been less commonly reported in the literature as compared to other GI symptoms, we unexpectedly observed that Twitter activity on the topic of pancreatology was the highest among all gastroenterology subtopics. However, there was no statistically significant correlation between Twitter and publication activity or between Twitter and preprint activity on the topic of pancreatology, thus implying that the rise of social media activity did not necessarily reflect the same rise in activity in bioRxiv, medRxiv, or PubMed activity. This could possibly be explained by reviewing the stakeholder data for pancreatology-related tweets. For each of the other gastroenterology subfields, stakeholder demographics fell largely under the categories of physicians, nonmedical doctors, or researchers and academic users; however, for the topic of “pancreas,” the most active stakeholders appeared to be advocacy organizations or patient advocates. Upon further review of the top 50 associated hashtags used for the “pancreas” topic, those related to cystic fibrosis support and awareness groups were the most common. Patient advocacy organizations not only play an important part in espousing awareness of specific diseases and patient populations, but they also serve an essential role in the dissemination of information on their behalf, typically in the form or raising awareness on the internet, lobbying directly for change within the government of other institutions, and via marketing and outreach, but likely only indirectly in the context of research activity. Therefore, this increase in Twitter activity may not translate to a proportionate increase in research publication activity on the topic of COVID-19 and pancreatology as demonstrated in this study. Using this topic as an example, further investigation diving deeper into the content of tweets, specific users, and how they translate directly into published work would be helpful to document the direct impact social media has on research activity.

Limitations

There are several limitations to this study that are worth noting. The first major limitation is that social media platforms, such as Twitter, are not available in several countries, including China. This could help to explain China’s lead in research publication activity as compared with other regions, as publication is likely one of the primary modalities used in this country for disseminating information. Other social media platforms, including WeChat and Sina Weibo, are used in China; however, information regarding the use of these other social media platforms within China is limited to date. Future studies are needed to assess social media activity on these alternative platforms and their association with publications as well as how they compared to the use of Twitter in other nations, such as the United States (where Twitter is not banned). Second, although we were able to account for and reassign duplicate publications for the various categorizations performed, a similar approach could not be guaranteed for Twitter data. The very nature of tweets allows for other users to publish an original tweet to their account generating additional impressions as it is sent to their followers (also known as a “retweet”). Therefore, limiting duplicate tweets may artificially decrease activity. More importantly, certain tweets, or even retweets, may have been assigned to more than one topic area owing to limitations of the Symplur software. Hence, we were unable to limit tweets to a single topic, and as such, this may have artificially boosted the overall number in certain organ system topics, thereby potentially skewing the results of this study.

Conclusions

In conclusion, this study demonstrates patterns in the utility of social media and publications—both in preprint repositories as well as peer-reviewed journals—for the rapid dissemination of information during a global pandemic of infectious disease. As the world faces this unprecedented public health emergency, this study has reflected on shifting worldwide trends from solely traditional methods of disseminating information (ie, via publications) to more contemporary methods, specifically among the GI community. social media tools like Twitter can be an effective method for educating and informing audiences in real time and via an interactive approach, a feat that cannot always be achieved with more conventional methods (ie, scientific publications). The new media age has resulted in a number of novel avenues for the distribution of information, including Twitter. Utilizing a single modality for dissemination of health care discoveries or information has been shown by the AHRQ to not be as effective as utilizing multiple modalities. It may therefore be time for the medical and scientific communities to cultivate formal social media platforms as effective tools for data sharing and collaboration to augment existing modalities of archival publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms used for Twitter, PubMed, and preprint repositories.

[DOCX File , 18 KB - jmir_v23i6e26956_app1.docx]

Multimedia Appendix 2

Supplementary tables summarizing additional information pertaining to COVID-19–related Twitter, publication, and preprint activity by organ system per half-month time intervals.

[DOCX File , 40 KB - jmir_v23i6e26956_app2.docx]

References

1. Coronavirus Disease (COVID-19)Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-03-03]
2. Zhang J, Dong X, Cao Y, Yuan Y, Yang Y, Yan Y, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy 2020 Jul;75(7):1730-1741. [doi: 10.1111/all.14238] [Medline: 32077115]
3. Mao R, Liang J, Shen J, Ghosh S, Zhu L, Yang H, Chinese Society of IBD, Chinese Elite IBD Union, Chinese IBD Quality Care Evaluation Center Committee. Implications of COVID-19 for patients with pre-existing digestive diseases. Lancet Gastroenterol Hepatol 2020 May;5(5):425-427 [FREE Full text] [doi: 10.1016/S2468-1253(20)30076-5] [Medline: 32171057]
4. Ding S, Liang TJ. Is SARS-CoV-2 also an enteric pathogen with potential fecal-oral transmission? a COVID-19 virological and clinical review. Gastroenterology 2020 Jul;159(1):53-61 [FREE Full text] [doi: 10.1053/j.gastro.2020.04.052] [Medline: 32353371]
5. Fan Z, Chen L, Li J, Cheng X, Yang J, Tian C, et al. Clinical features of COVID-19-related liver functional abnormality. Clin Gastroenterol Hepatol 2020 Jun;18(7):1561-1566 [FREE Full text] [doi: 10.1016/j.cgh.2020.04.002] [Medline: 32283325]
6. Donato MF, Invernizzi F, Lampertico P, Rossi G. Health status of patients who underwent liver transplantation during the coronavirus outbreak at a large center in Milan, Italy. Clin Gastroenterol Hepatol 2020 Aug;18(9):2131-2133.e1 [FREE Full text] [doi: 10.1016/j.cgh.2020.04.041] [Medline: 32334081]
7. D'Amico F, Peyrin-Biroulet L, Danese S. Inflammatory bowel diseases and COVID-19: the invisible enemy. Gastroenterology 2020 Jun;158(8):2302-2304 [FREE Full text] [doi: 10.1053/j.gastro.2020.04.032] [Medline: 32305331]
8. Rubin DT, Abreu MT, Rai V, Siegel CA, International Organization for the Study of Inflammatory Bowel Disease. Management of patients with Crohn's disease and ulcerative colitis during the coronavirus disease-2019 pandemic: results of an international meeting. Gastroenterology 2020 Jul;159(1):6-13.e6 [FREE Full text] [doi: 10.1053/j.gastro.2020.04.002] [Medline: 32272113]
9. Rodríguez-Lago I, Ramírez de la Piscina P, Elorza A, Merino O, Ortiz de Zárate J, Cabriada JL. Characteristics and prognosis of patients with inflammatory bowel disease during the SARS-CoV-2 pandemic in the Basque Country (Spain). Gastroenterology 2020 Aug;159(2):781-783 [FREE Full text] [doi: 10.1053/j.gastro.2020.04.043] [Medline: 32330477]
10. Repici A, Maselli R, Colombo M, Gabbiadini R, Spadaccini M, Anderloni A, et al. Coronavirus (COVID-19) outbreak: what the department of endoscopy should know. Gastrointest Endosc 2020 Jul;92(1):192-197 [FREE Full text] [doi: 10.1016/j.gie.2020.03.019] [Medline: 32179106]

11. Cennamo V, Bassi M, Landi S, Apolito P, Ghersi S, Dabizzi E, et al. Redesign of a GI endoscopy unit during the COVID-19 emergency: a practical model. Dig Liver Dis 2020 Oct;52(10):1178-1187 [FREE Full text] [doi: 10.1016/j.dld.2020.05.007] [Medline: 32425734]
12. Dunford D, Dale B, Stylianou N, Lowther E, Ahmed M, de la Torre Arenas I. Coronavirus: The world in lockdown in maps and charts. BBC News. 2020 Apr 7. URL: https://www.bbc.com/news/world-52103747 [accessed 2021-04-12]
13. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):taaa031 [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
14. Merchant RM, Lurie N. Social media and emergency preparedness in response to novel coronavirus. JAMA 2020 May 26;323(20):2011-2012. [doi: 10.1001/jama.2020.4469] [Medline: 32202611]
15. Pershad Y, Hangge PT, Albadawi H, Oklu R. Social medicine: Twitter in healthcare. J Clin Med 2018 May 28;7(6):121. [doi: 10.3390/jcm7060121] [Medline: 29843360]
16. Merchant RM. Evaluating the potential role of social media in preventive health care. JAMA 2020 Feb 04;323(5):411-412. [doi: 10.1001/jama.2019.21084] [Medline: 31922532]
17. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
18. Brady RRW, Chapman SJ, Atallah S, Chand M, Mayol J, Lacy AM, et al. #colorectalsurgery. Br J Surg 2017 Oct;104(11):1470-1476. [doi: 10.1002/bjs.10615] [Medline: 28881004]
19. Symplur Signals. URL: https://www.symplur.com/products/signals/ [accessed 2021-05-24]
20. McCormack L, Sheridan S, Lewis M, Boudewyns V, Melvin CL, Kistler C, et al. Communication and dissemination strategies to facilitate the use of health-related evidence. Evid Rep Technol Assess (Full Rep) 2013 Nov(213):1-520. [doi: 10.23970/ahrqepcerta213] [Medline: 24423078]
21. Cohen J, Normile D. New SARS-like virus in China triggers alarm. Science 2020 Jan 17;367(6475):234-235. [doi: 10.1126/science.367.6475.234] [Medline: 31949058]
22. Teixeira da Silva JA, Tsigaris P, Erfanmanesh M. Publishing volumes in major databases related to Covid-19. Scientometrics 2020 Aug 28:1-12 [FREE Full text] [doi: 10.1007/s11192-020-03675-3] [Medline: 32904414]
23. Fox CS, Bonaca MA, Ryan JJ, Massaro JM, Barry K, Loscalzo J. A randomized trial of social media from Circulation. Circulation 2015 Jan 06;131(1):28-33. [doi: 10.1161/CIRCULATIONAHA.114.013509] [Medline: 25406308]
24. Fox CS, Gurary EB, Ryan J, Bonaca M, Barry K, Loscalzo J, et al. Randomized controlled trial of social media: effect of increased intensity of the intervention. J Am Heart Assoc 2016 Apr 27;5(5):e003088 [FREE Full text] [doi: 10.1161/JAHA.115.003088] [Medline: 27121850]
25. Eysenbach G. Can tweets predict citations? metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 2011 Dec 19;13(4):e123 [FREE Full text] [doi: 10.2196/jmir.2012] [Medline: 22173204]
26. Bardus M, El Rassi R, Chahrour M, Akl EW, Raslan AS, Meho LI, et al. The use of social media to increase the impact of health research: systematic review. J Med Internet Res 2020 Jul 06;22(7):e15607 [FREE Full text] [doi: 10.2196/15607] [Medline: 32628113]

Abbreviations

ACE2: angiotensin-2
AHRQ: Agency for Healthcare Research and Quality
GI: gastrointestinal
IBD: inflammatory bowel disease
MERS: Middle East Respiratory Syndrome
SARS: severe acute respiratory syndrome coronavirus

Edited by C Basch; submitted 05.01.21; peer-reviewed by C Hudak, S Mathews, K Reuter; comments to author 11.02.21; revised version received 05.05.21; accepted 17.05.21; published 17.06.21.

Please cite as:
Taneja SL, Passi M, Bhattacharya S, Schueler SA, Gurram S, Koh C
Social Media and Research Publication Activity During Early Stages of the COVID-19 Pandemic: Longitudinal Trend Analysis
J Med Internet Res 2021;23(6):e26956
URL: https://www.jmir.org/2021/6/e26956
doi:10.2196/26956
PMID:33974550

Original Paper

Language and Sentiment Regarding Telemedicine and COVID-19 on Twitter: Longitudinal Infodemiology Study

Catherine C Pollack[1,2], BS; Diane Gilbert-Diamond[2,3,4], ScD; Jennifer A Alford-Teaster[1,2], MA, MPH; Tracy Onega[5], PhD

[1]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States
[2]Department of Epidemiology, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States
[3]Department of Pediatrics, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States
[4]Department of Medicine, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States
[5]Department of Population Health Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Catherine C Pollack, BS
Department of Biomedical Data Science
Geisel School of Medicine at Dartmouth College
1 Medical Center Drive
Lebanon, NH, 03766
United States
Phone: 1 540 497 3419
Email: Catherine.c.pollack.gr@dartmouth.edu

Abstract

Background: The COVID-19 pandemic has necessitated a rapid shift in how individuals interact with and receive fundamental services, including health care. Although telemedicine is not a novel technology, previous studies have offered mixed opinions surrounding its utilization. However, there exists a dearth of research on how these opinions have evolved over the course of the current pandemic.

Objective: This study aims to evaluate how the language and sentiment surrounding telemedicine has evolved throughout the COVID-19 pandemic.

Methods: Tweets published between January 1, 2020, and April 24, 2021, containing at least one telemedicine-related and one COVID-19–related search term (“telemedicine-COVID”) were collected from the Twitter full archive search (N=351,718). A comparator sample containing only COVID-19 terms (“general-COVID”) was collected and sampled based on the daily distribution of telemedicine-COVID tweets. In addition to analyses of retweets and favorites, sentiment analysis was performed on both data sets in aggregate and within a subset of tweets receiving the top 100 most and least retweets.

Results: Telemedicine gained prominence during the early stages of the pandemic (ie, March through May 2020) before leveling off and reaching a steady state from June 2020 onward. Telemedicine-COVID tweets had a 21% lower average number of retweets than general-COVID tweets (incidence rate ratio 0.79, 95% CI 0.63-0.99; *P*=.04), but there was no difference in favorites. A majority of telemedicine-COVID tweets (180,295/351,718, 51.3%) were characterized as “positive,” compared to only 38.5% (135,434/351,401) of general-COVID tweets (*P*<.001). This trend was also true on a monthly level from March 2020 through April 2021. The most retweeted posts in both telemedicine-COVID and general-COVID data sets were authored by journalists and politicians. Whereas the majority of the most retweeted posts within the telemedicine-COVID data set were positive (55/101, 54.5%), a plurality of the most retweeted posts within the general-COVID data set were negative (44/89, 49.4%; *P*=.01).

Conclusions: During the COVID-19 pandemic, opinions surrounding telemedicine evolved to become more positive, especially when compared to the larger pool of COVID-19–related tweets. Decision makers should capitalize on these shifting public opinions to invest in telemedicine infrastructure and ensure its accessibility and success in a postpandemic world.

(J Med Internet Res 2021;23(6):e28648) doi:10.2196/28648

KEYWORDS

telemedicine; telehealth; COVID-19 pandemic; social media; sentiment analysis; Twitter; COVID-19; pandemic

Introduction

The COVID-19 public health crisis has transformed how individuals interact with critical services. This is particularly true of health care systems, which have been overwhelmed by patients with COVID-19 in both inpatient and intensive care units [1]. The sudden disruption in the ability to receive medical care has had widespread consequences for millions of Americans, and a survey by the US Centers for Disease Control and Prevention estimated that 40.9% of adults have delayed receiving medical care (including both emergency and routine care) due to concerns surrounding COVID-19 [2]. The new burdens placed on health care systems by the global pandemic have demonstrated the urgent need for the implementation of technologies to facilitate enhanced connectivity between patients and providers.

Telemedicine, defined as the delivery of health care services through electronic, audiovisual telecommunication systems, is not a novel concept—in fact, it has proven to be successful across a myriad of health domains [3]. The increased use of telemedicine during an emergency situation is also not new, and an increased uptake of telemedicine technologies has been observed during local, national, and international crises [4]. The implementation of telemedicine during a global pandemic is of particular value—not only can it be used to screen, diagnose, and triage patients from the comfort of their own homes, but it can also limit the need for physicians to use personal protective equipment that may be in low supply, facilitate rapid follow-up with diverse patient populations (particularly older patients or those living in a rural environment), reduce exposure to the infectious agent, and decrease the risk for intrahospital infection [5,6].

Despite its promise, prepandemic uptake of telemedicine in the United States was limited largely due to a lack of physician acceptance, stringent and heterogeneous licensing and reimbursement policies, and the upfront monetary costs of investing in the necessary infrastructure [4,5]. Nevertheless, shifts in policies in the peripandemic period have led to a surge in the uptake of telemedicine technologies. After the US Centers for Medicare & Medicaid Services revised their telemedicine reimbursement policies to include over 135 services (including emergency department visits, inpatient and nursing facility visits, and "discharge day management services"), over 36% of Medicare beneficiaries received at least one telemedicine service. Similarly, Medicaid and the Children's Health Insurance Program reported a 2600% increase in the use of telemedicine services in March through June 2020 compared to the same period in 2019 [7]. Thus, there exists a clear demand for heterogeneous patient populations to receive care digitally. However, these utilization metrics may not directly correlate to patient satisfaction, willingness to engage in telemedicine, or their ability to do so. Thus, additional work is needed to quantify patients' perceptions of the enhanced accessibility of telemedicine services.

Social media has increasingly been used as a surveillance tool by public health researchers to answer diverse health-related questions, including detecting disease outbreaks, situational awareness of humanitarian crises (such as natural disasters), and understanding a population's reaction towards certain messaging or events [8]. This is particularly vital during an ongoing public health pandemic, wherein social media can provide insights more rapidly than traditional data collection methodologies such as surveys [9]. One of the more commonly used social media platforms for this purpose is Twitter, owing to the abundance of daily content and the widespread (although not necessarily representative) demographic reach of this platform [10]. A previous study evaluating telemedicine discourse on Twitter during the COVID-19 pandemic has identified distinct user networks that bridge content domains and user types (including educational, promotional, and political materials) [11], whereas another study found that the geographic distribution of telemedicine tweets in the United States was significantly correlated to the number of confirmed COVID-19 cases within a state [12]. However, both studies only evaluated a 1- to 2-week period, leaving a gap in understanding how conversations surrounding telemedicine may have changed over the course of the pandemic.

Furthermore, neither study considered the sentiment content of the tweets, which could provide more precise insights into how Twitter users perceive telemedicine. A separate evaluation of the sentiment of telemedicine-related tweets within a subpopulation of health care providers found overall positive opinions that focused on safety, accessibility, and implementation strategies [13]. However, because this study only focused on providers, there exists a dearth of research on the sentiment toward telemedicine expressed by the general Twitter population. Thus, the purpose of this study was to characterize how the content of telemedicine-related tweets has evolved during the COVID-19 pandemic, with a particular focus on changes in sentiment types. It was hypothesized that the frequency of telemedicine-related content on Twitter increased throughout the COVID-19 pandemic and that the sentiment of tweets became more positive across this period.

Methods

Data Collection

The characterization of telemedicine content in relation to COVID-19 was evaluated across 4 components of tweets: the number of favorites and retweets, language used within the tweet, sentiment of the tweet, and authorship. Tweets in English language posted between January 1, 2020, and April 24, 2021, containing at least one of a series of telemedicine-related terms and at least one of a series of COVID-19–related terms were curated from the Twitter full-archive search available via the Academic Research product track [14]. Search terms were derived from the literature and the Medical Subject Headings thesaurus established by the National Library of Medicine (NLM) of the National Institutes of Health (NIH) (Multimedia Appendix 1) [11,15]. A random sample of tweets containing only COVID-19–related terms was extracted for the same period for comparison with the distribution of sampled tweets matched with the daily distribution of telemedicine-related COVID-19 tweets. No geographic restrictions regarding the location of the tweet were implemented, and all tweets beginning with “RT @” (indicating retweet status) were removed prior to analysis.

Text Processing

Standard natural language processing preprocessing procedures were performed prior to textual analysis of the tweets. First, links, mentions, hashtags, and HTML escape characters were removed from all tweets. Next, tweet-level sentiment was calculated using the Valence Aware Dictionary and sEntiment Reasoner (VADER), which was designed specifically for use on a social media corpus and has been validated in other Twitter-based studies, including one on telemedicine [16,17]. In addition to the compound sentiment score, a categorical sentiment was assigned to each tweet based on prespecified cutoffs within the literature (ie, *positive* if the compound sentiment was greater than 0.05, *negative* if the compound sentiment was lesser than –0.05, and *neutral* if the compound sentiment was between these values) [16]. After determining the sentiment, other preprocessing included tokenizing and lemmatizing the text and removing traditional English and Spanish stop words. In addition, words with an inverse document frequency in the 0.05th percentile were removed given their high frequency across all tweets (eg, “covid19”). Other words removed included non-English words (as determined by the GradyAugmented dataset [18]), as well as words with alternative connotations that may have skewed any analysis (eg, “trump” could refer to the verb or the 45th President of the United States; “patient” could refer to the adjective or a person receiving medical treatment). This processing was performed for both single words as well as bigrams (ie, two-word phrases) and trigrams (ie, three-word phrases).

Author Analysis

In addition to evaluating all telemedicine-related COVID-19 tweets (“telemedicine-COVID”) relative to a general COVID-19 sample (“general-COVID”), the authors of tweets with the 100 most and the 100 fewest retweets were extracted and manually labeled with a domain (eg, “news,” “political,” “health and medicine”) and account type (ie, “organization” or “individual”). Domains were created by reviewing the author description information and, for verified users, confirming their identity through an independent Google search. Individual authors who represented a nonverified person were automatically labeled as “private citizen,” whereas all organizations were labeled with a domain regardless of the verification status. This process was completed by one member of the research team for both the telemedicine-COVID and general-COVID data sets to compare outcomes within these tweet subpopulations. A second member of the research team independently labeled 10% of the authors as a “validity check” [19]. This subsequentially resulted in Cohen κ=0.66 (percent agreement: 77.3%) for domain type and κ= 0.75 (percent agreement: 86.4%) for the account type.

Statistical Analysis

A comparative analysis of retweets and favorites by month and by data set (ie, telemedicine-COVID vs general-COVID) was conducted using zero-inflated Poisson regression to accommodate for the preponderance of tweets receiving no favorites or retweets and the count nature of the outcome variable. Sentiment analysis was also performed to assess the distribution of positive, negative, and neutral tweets both overall and by month and compared using chi-square tests. This analysis was repeated to compare sentiment in the telemedicine-COVID data set with that in the general-COVID data set. Chi-square tests were also used to compare the distribution of sentiment between the top 100 most and least retweeted posts both within each data set and between them. All analyses were conducted in Python (version 3.7.4) within the Jupyter Notebook graphical user interface (GUI) (version 7.19.0) and R (version 3.6.3) within the RStudio GUI (version 1.3.959). Code is available on GitHub [20]; tweet IDs are available upon request. Given the public nature of social media data, institutional review board approval was not required as specified in Regulation 45 CFR 46 as authored by the US Department of Health and Human Services Office for Human Research Protections [21].

Results

After removing duplicate tweets and retweets from analysis, 351,718 tweets related to telemedicine and COVID-19 were used in the analytic data set (telemedicine-COVID). A sample comparator data set (general-COVID) of 351,401 tweets was collected with a similar daily distribution for comparison.

Telemedicine-COVID tweets spiked in the early stages of the pandemic (eg, March 2020 to May 2020) before leveling off in the following months (Figure 1). There were significant monthly variations in the number of retweets of telemedicine-COVID tweets (Figure 2). Retweets peaked in March 2020, with April, May, July, September, October, and December 2020 and January through April 2021 having significantly fewer retweets by comparison (Table S1 in Multimedia Appendix 2). Tweets from the telemedicine-COVID data set had a 21% lower average number of retweets than tweets from the general-COVID data set (incidence rate ratio [IRR] 0.79, 95% CI 0.63-0.99; *P*=.04; Table S2 in Multimedia Appendix 2). In terms of favorites, telemedicine-COVID tweets in April, May, July, August, September, October, and December 2020, and tweets from January through April 2021 had a significantly lower average number of favorites than those in March 2020 (Table S3 in Multimedia Appendix 2). Telemedicine-COVID tweets had a 14% lower average number of favorites than did general-COVID tweets, but this value was not statistically significant (IRR 0.86, 95% CI 0.58-1.28; *P*=.45) (Table S4 in Multimedia Appendix 2).

When broken down by sentiment, the percentage of telemedicine-COVID tweets with an overall sentiment of positive or neutral generally increased over the period of interest, whereas the percentage of negative tweets decreased (Figure 3). There were significant monthly variations in the percentage of tweets with each sentiment (*P*<.001). The months with the highest percentage of positive telemedicine-COVID tweets were August 2020 (14,371/24,543, 58.6%), September 2020 (10,475/18,758, 55.8%), and March 2020 (19,851/36,478, 54.4%), whereas the months with the highest percentage of negative tweets were February 2020 (453/1274, 35.6%), January 2020 (13/43, 30.2%), and January 2021 (3695/16,613, 22.2%). There were also significant differences in the distribution of positive, negative, and neutral tweets between the telemedicine-COVID and general-COVID data sets (*P*<.001; Figure 4). Although sentiments were evenly distributed among general-COVID tweets (38.5% positive, 31.4% neutral, and 30.0% negative), a majority of telemedicine-COVID tweets (N=351,718) were positive (n=180,295, 51.3%), followed by neutral (n=100,870, 28.7%). There were also significant variations between data sets on a month-to-month basis for March 2020 through April 2021, with a significantly higher proportion of positive tweets in the telemedicine-COVID data set relative to the general-COVID data set (Multimedia Appendix 3).

Figure 1. Frequency of tweets mentioning telemedicine and COVID-19 between January 1, 2020, and April 24, 2021 (N=351,718).

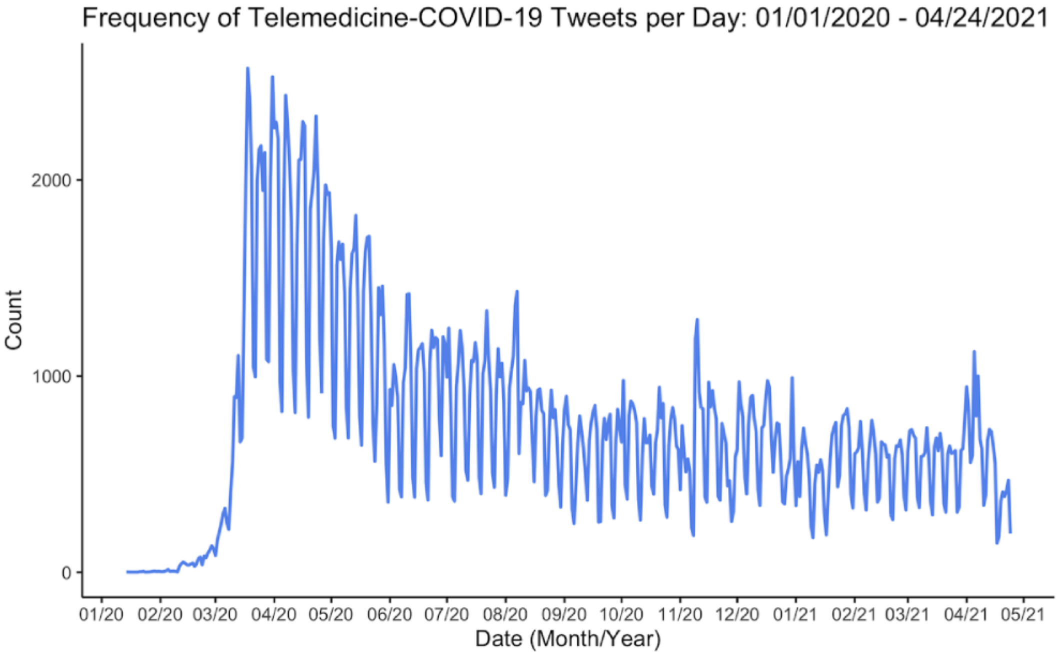


Figure 2. Monthly variations in the base 10 log number of retweets for tweets mentioning COVID-19 and telemedicine. Note that 219,212 tweets (62.2%) had no retweets and are not included in this visualization.

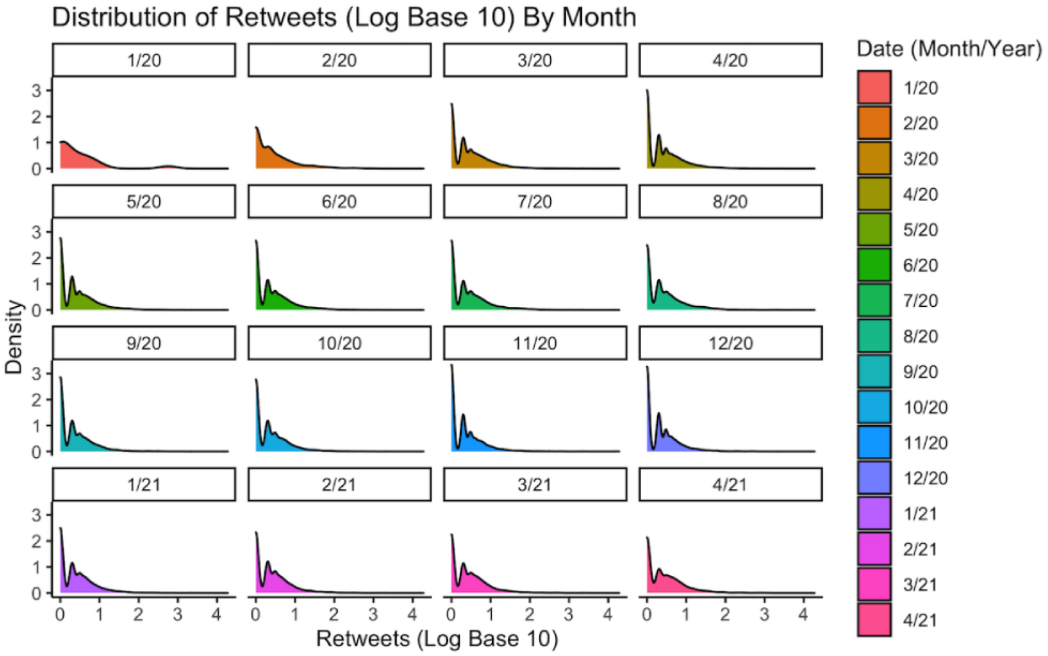


Figure 3. Changes in the frequency of positive, negative, and neutral tweets mentioning telemedicine and COVID-19 posted between January 1, 2020, and April 24, 2021 (N=351,718).

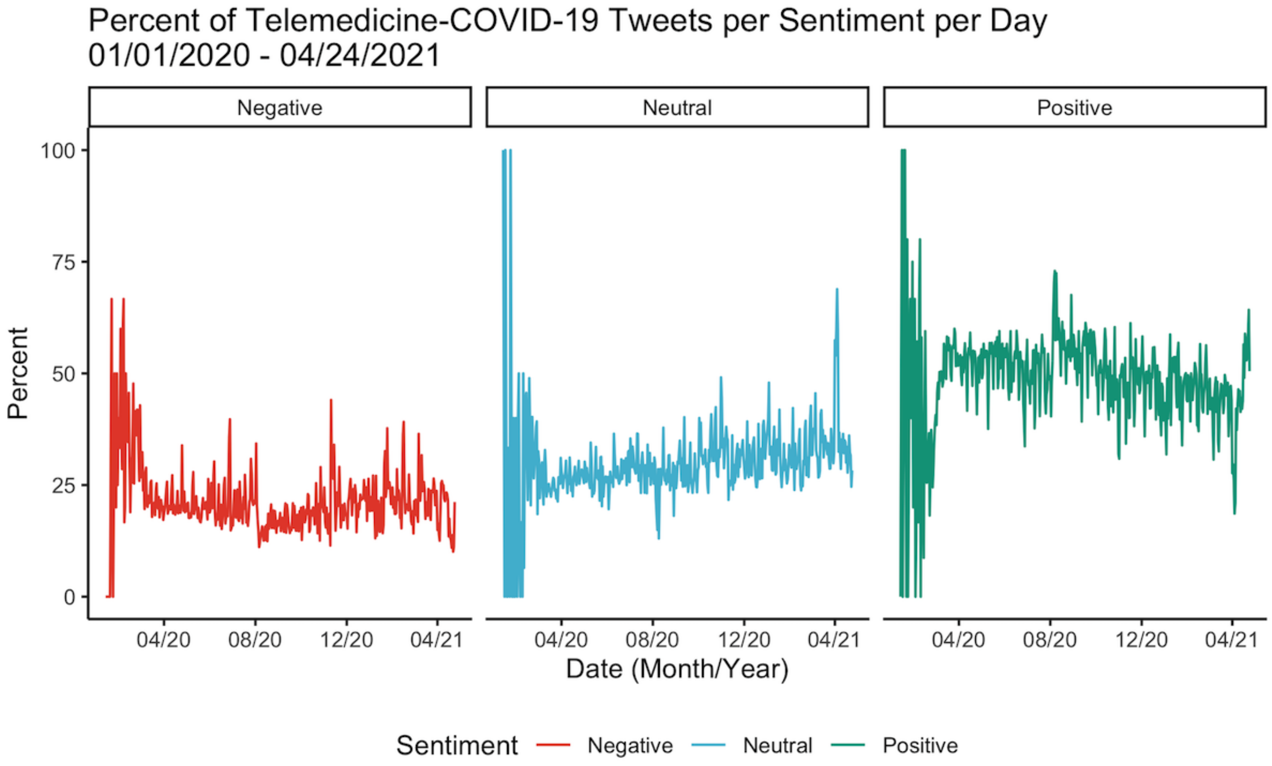
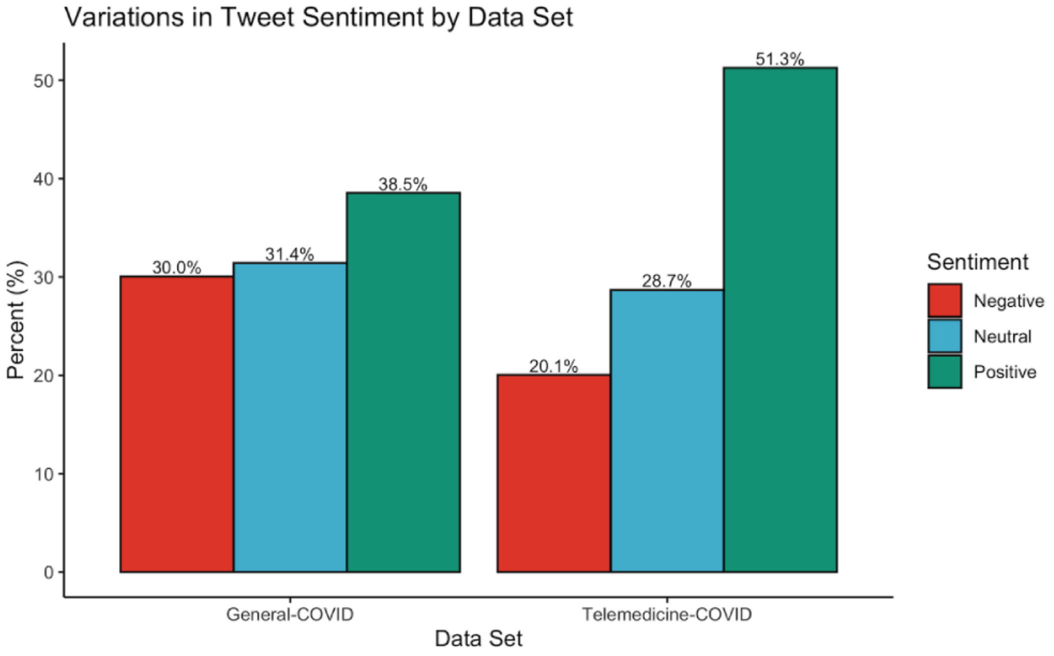


Figure 4. Variations in sentiment by tweet type (ie, telemedicine and COVID-19 vs COVID-19 only). A significantly higher proportion of tweets with a positive sentiment was found in the telemedicine-COVID data set than in the general-COVID data set (*P*<.001).



Within the telemedicine-COVID tweet data set (N=351,718), the most frequently used bigrams were “town hall” (n=1411, 0.40%); “white house” (n=848, 0.24%), “san diego” (n=757, 0.22%), “mask wearing” (n=670, 0.19%), “lessons learned” (n=610, 0.17%), and “artificial intelligence” (n=562, 0.16%). The most frequently used trigrams were “widespread mask wearing” (n=385, 0.11%), “feeling helpless hopeless” (n=301, 0.09%), “fast track vital” (n=199, 0.06%), “cancer sooner simply” (n=191, 0.05%), “thermal imaging cameras” (n=188, 0.05%), and “list refresh page” (n=176, 0.05%).

In contrast, within the general-COVID sample (N=351,401), the most frequently used bigrams were “ill [*sic*] deliver” (n=1195, 0.34%), “nursing homes” (n=693, 0.20%), “task force” (n=597, 0.17%), “herd immunity” (n=563, 0.16%), “prime minister” (n=508, 0.14%), and “town hall” (n=439, 0.12%).

The most frequently used trigrams were “appointment detected provider” (n=214, 0.06%), “personal protective equipment” (n=205, 0.06%), “complete project chapter” (n=181, 0.05%), “wall street journal” (n=101, 0.03%), “operation warp speed” (n=68, 0.02%), and “midnight mm rain” (n=67, 0.02%). A word cloud of unigrams is presented in Multimedia Appendix 4.

When broken down by authorship, the top 100 most retweeted telemedicine-COVID tweets were predominantly authored by politicians (20/93, 21.5%) and private citizens (20/93, 21.5%), followed by journalists (14/93, 15.1%). Similarly, the top 100 most retweeted general-COVID sample tweets were predominantly authored by private citizens (23/89, 25.8%), journalists (16/89, 18.0%), and politicians (13/89, 14.6%). Of the sample of tweets without any retweets, 82.8% (72/87) of those from the general-COVID sample and 61.1% (55/90) from the telemedicine-COVID sample were authored by private citizens. There was a small subset of authors within each data set that did not have available account information (7 authors of the top 100 telemedicine-COVID retweets, 10 authors of the bottom 100 telemedicine-COVID retweets, 11 authors of the top 100 general-COVID retweets, and 13 authors of the bottom 100 general-COVID retweets). In terms of sentiment, of the most retweeted tweets within the telemedicine-COVID data set, 54.5% (55/101) were positive and 30.7% (31/101) were negative. In contrast, of the fewest retweeted tweets, 51.3% (40/78) were positive and 19.2% (15/78) were negative (*P*=.03). Within the general-COVID data set, 49.4% (44/89) of the top retweeted tweets were negative and 33.7% (30/89) were positive. This was not significantly different from the fewest retweeted general-COVID posts (*P*=.16), but it was significantly different from the most retweeted telemedicine-COVID tweets (*P*=.01). The sentiment of the fewest retweeted general-COVID did not significantly differ from the sentiment of the fewest retweeted telemedicine-COVID tweets (*P*=.07).

Discussion

Principal Findings

To the best of our knowledge, this study is the first to comprehensively evaluate tweets pertaining to telemedicine and COVID-19 posted between January 2020 and April 2021 compared to a general COVID-19 data set. Analysis of retweets and favorites suggested monthly variations in the “attention” received by telemedicine-COVID tweets, and these tweets had a significantly lower average number of retweets than the general-COVID data set. Telemedicine-COVID tweets were predominantly positive both overall and by month, especially compared to the general-COVID data set. There were also variations in the frequency and sentiment of tweets made by entities in various domains, including private citizens, politicians, and journalists, and a higher proportion of the most retweeted tweets in the telemedicine-COVID data set were positive than those in the general-COVID data set. The findings presented here demonstrate how social media can be leveraged to perform surveillance of shifting opinions surrounding critical health technologies, including telemedicine.

The number of tweets that mentioned telemedicine and COVID-19 drastically increased between February and March 2020, coinciding with the declaration of the COVID-19 pandemic by the World Health Organization on March 11, 2020 [22]. As stay-at-home orders continued throughout April and May 2020, telemedicine became more prominent within the COVID-19 dialogue on Twitter, and health systems began to adapt to the use of this technology. Discussions surrounding telemedicine decreased through June 2020 and remained relatively constant thereafter, perhaps reflecting the achievement of “steady state.” This leveling off coincides with previous studies that have found that the weekly rate of telemedicine consults in a Medicare population peaked in April before declining in June 2020 [23].

As the volume of telemedicine-COVID tweets evolved, so did the sentiment of these tweets. Over half of all months between March 2020 and April 2021 had a majority of telemedicine-COVID tweets labeled as positive, and all months during this period had a significantly higher proportion of positive tweets compared to the general-COVID data set. Although this finding aligns with a study prior to the pandemic that found a higher number of positive telemedicine tweets than negative and neutral tweets, it contrasts with another study that found that 59% of individuals were either unsure of telemedicine or considered it subpar to traditional care delivery mechanisms [17,24]. The positive results found in this study may be reflective of increasing acceptance toward telemedicine—while patients may have viewed telemedicine as just an alternative to in-person care before the pandemic, telemedicine may now be viewed as an alternative to no care whatsoever. This finding aligns with recent patient surveys on their opinion of telemedicine, which found that 79% of patients were “satisfied” with their experience with telemedicine and 78% felt that they had a health problem that could be addressed virtually [25]. Furthermore, increased positivity in the middle stages of the pandemic (particularly in August and September 2020) reflects the evolving understanding of the Twitter population that reduced contact during a telemedicine visit can provide a safer experience during a highly contagious disease outbreak. The increase in positivity in this period in particular also corresponds to the signing of the Executive Order 13941 on August 3, 2020, which aimed to “improve rural health and telehealth access” to Medicare beneficiaries during the postpandemic period [26,27].

The most followed accounts within both the telemedicine-COVID and general-COVID data sets predominantly consisted of journalists and politicians. These groups present stark contrasts in content veracity—journalists likely provide neutral content with minimal bias (depending on the agency), whereas politicians likely share more polarized content that reflect their own views. Prior work evaluating tweets on telemedicine during the COVID-19 pandemic (but not necessarily related to it) found that private citizens frequently retweeted content from both sources, although politicians were retweeted more frequently [11]. Thus, there is a chance that a larger number of Twitter users’ opinions on telemedicine and COVID-19 may be limited to “echo chambers” that reinforce their own opinion and, in a worst-case scenario, spread misinformation with deleterious consequences. However, the fact that a majority of tweets within the telemedicine-COVID data set were positive suggests that this may not necessarily be the case for tweets pertaining to telemedicine or other novel technologies. Conjecturally, this may reflect that positive news and experiences on telemedicine were amplified during the COVID-19 pandemic, which could lead to more widespread adoption and uptake of this technology.

Policy Implications

Telemedicine is the pinnacle "21st century approach" to deliver convenient and less expensive care, and over 50 large US health systems have integrated it into their standard operating procedures [5]. Although the surge in telemedicine visits during the early phases of the pandemic have waned to some extent, evidence shows a high (but geographically variable) degree of persistence of telemedicine, from 8% to almost 48% [23]. This can likely be attributed to relaxations in policies that presented major challenges to telemedicine uptake, including reimbursement parity, interstate licensing, prescribing practices, the use of Health Insurance Portability and Accountability Act (HIPAA)-compliant technologies, and the definition of an "originating site" (ie, where the patient is located), to name a few [28,29]. The overwhelming and consistent positive nature of conversations surrounding telemedicine on Twitter as presented here, coupled with a surge in uptake throughout the pandemic, demonstrate that diverse, heterogeneous populations view telemedicine favorably, including patients, providers, and politicians. These findings provide clear evidence for policymakers that replacing restrictive policies with long-term, systematic favorable toward telemedicine would be met with support from numerous, diverse communities.

Limitations and Future Directions

Despite the promise of this study, it is not without its limitations. Although the list of search terms was comprehensive and based on prior literature, it may have missed tweets that used other terminology to describe telemedicine-related services. Similarly, it is possible that some of the included tweets may not have directly been related to telemedicine. The analysis also does not include telemedicine-related tweets that were made during the study period that did not contain a direct mention of COVID-19. Although this ensured that the data specifically focused on the impacts of telemedicine as they related to COVID-19, future work could analyze how telemedicine tweets that directly mentioned COVID-19 varied from those that did not during the same time. Furthermore, future work could evaluate whether the trends observed in the varied from tweets made during the same period in the prior year. In addition, the present work does not include geospatial data, which has previously been shown to be an influential component of tweet sentiment [30,31]. Thus, future work could evaluate geospatial variations in telemedicine sentiment, including whether it is associated with uptake of the technology in local medical facilities. In addition, only the top 100 most and least retweeted posts were included in the author-level analysis, and future work could expand on this to label more accounts, improve the classification of labels (eg, labeling fewer accounts as "private citizens"), or analyze only a subset of these accounts (eg, politicians or "influencers"). Lastly, the Twitter population of predominantly 18-24 years old, well-educated individuals is not representative of a US or broader population [32]. Thus, future work is needed to characterize the change in sentiment within other populations, especially those that may not be technologically literate and may therefore encounter barriers when attempting to utilize telemedicine.

Conclusions

Opinions on telemedicine and COVID-19 on Twitter have increased in popularity and were largely positive throughout 2020 and the beginning of 2021. These telemedicine-COVID tweets were generally more positive than general-COVID tweets both overall and within the subset of the most followed authors, suggesting an amplification of discussion surrounding the benefits of telemedicine. Given the relative positivity with which individuals seemed to view telemedicine during the COVID-19 pandemic, shifts in policies stemming from the COVID-19 pandemic that support telemedicine are likely to be well received.

Acknowledgments

This publication was supported in part by an NLM-funded predoctoral fellowship to CCP (5T32LM012204-05). This work was also supported by National Cancer Institute (NCI) (grant P30CA023108-41S5). The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the NLM, NCI, or NIH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Telemedicine and COVID-19–related search terms used to select tweets.
[DOCX File , 14 KB - jmir_v23i6e28648_app1.docx]

Multimedia Appendix 2
Zero-inflated Poisson models.
[DOCX File , 26 KB - jmir_v23i6e28648_app2.docx]

Multimedia Appendix 3

Sentiment variations between the telemedicine-COVID data set and the general-COVID data set by month.
[DOCX File , 11512 KB - jmir_v23i6e28648_app3.docx]

Multimedia Appendix 4
Word clouds for unigrams for telemedicine-specific COVID-19 tweets and general COVID-19 sample tweets.
[DOCX File , 1704 KB - jmir_v23i6e28648_app4.docx]

References

1. Shammas B, Cha A, Guarino B, Dupree J. Record numbers of covid-19 patients push hospitals and staffs to the limit. Washington Post. 2020 Dec 16. URL: https://www.washingtonpost.com/health/2020/12/16/hospitals-covid-overwhelmed/ [accessed 2020-12-18]
2. Czeisler ME, Marynak K, Clarke KEN, Salah Z, Shakya I, Thierry JM, et al. Delay or avoidance of medical care because of COVID-19-related concerns - United States, June 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 11;69(36):1250-1257 [FREE Full text] [doi: 10.15585/mmwr.mm6936a4] [Medline: 32915166]
3. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
4. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Mar 20;26(5):309-313. [doi: 10.1177/1357633x20916567]
5. Hollander JE, Carr BG. Virtually perfect? telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
6. Bokolo AJ. Exploring the adoption of telemedicine and virtual software for care of outpatients during and after COVID-19 pandemic. Ir J Med Sci 2021 Feb;190(1):1-10 [FREE Full text] [doi: 10.1007/s11845-020-02299-z] [Medline: 32642981]
7. Trump administration drives telehealth services in Medicaid and Medicare. Centers for Medicare & Medicaid Services. 2020 Oct 14. URL: https://www.cms.gov/newsroom/press-releases/trump-administration-drives-telehealth-services-medicaid-and-medicare [accessed 2020-12-18]
8. Fung IC, Tse ZTH, Fu K. The use of social media in public health surveillance. Western Pac Surveill Response J 2015;6(2):3-6 [FREE Full text] [doi: 10.5365/WPSAR.2015.6.1.019] [Medline: 26306208]
9. Murray CJL, Alamro NMS, Hwang H, Lee U. Digital public health and COVID-19. Lancet Public Health 2020 Sep;5(9):e469-e470 [FREE Full text] [doi: 10.1016/S2468-2667(20)30187-0] [Medline: 32791051]
10. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
11. Lit CL, Khayal IS. Understanding Twitter telehealth communication during the COVID-19 pandemic using hetero-functional graph theory. 2020 Presented at: 2020 IEEE International Smart Cities Conference (ISC2); September 28, 2020 - October 1, 2020; Piscataway, New Jersey, USA p. 1-6. [doi: 10.1109/isc251055.2020.9239034]
12. Massaad E, Cherfan P. Social media data analytics on telehealth during the COVID-19 pandemic. Cureus 2020 Apr 26;12(4):e7838 [FREE Full text] [doi: 10.7759/cureus.7838] [Medline: 32467813]
13. Larson S, Popov V, Ali A, Ramanathan P, Jung S. Healthcare professionals' perceptions of telehealth: analysis of tweets from pre- and during the COVID-19 pandemic. In: Ruis AR, Lee SB, editors. Advances in Quantitative Ethnography. ICQE 2021. Communications in Computer and Information Science, vol 1312. Cham: Springer; Jan 22, 2021:390-405.
14. Academic research - Preparing for the application. Twitter Developer Platform. URL: https://developer.twitter.com/en/solutions/academic-research/application-info [accessed 2021-05-18]
15. MeSH (Medical Subject Headings). National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/mesh/ [accessed 2021-01-04]
16. Hutto CJ, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. 2014 Jun Presented at: Eight International Conference on Weblogs and Social Media (ICWSM-14); June 2, 2014 - June 4, 2014; Ann Arbor, Michigan, USA.
17. Talpada H, Halgamuge MN, Tran Quoc Vinh N. An analysis on use of deep learning and lexical-semantic based sentiment analysis method on Twitter data to understand the demographic trend of telemedicine. 2019 Presented at: 11th International Conference on Knowledge and Systems Engineering (KSE); October 24, 2019 - October 26, 2019; Da Nang, Vietnam p. 1-9. [doi: 10.1109/kse.2019.8919363]
18. Rinker T. GradyAugmented: augmented list of Grady Ward's English words and Mark Kantrowitz's names list. RDocumentation. URL: https://www.rdocumentation.org/packages/qdapDictionaries/versions/1.0.7/topics/GradyAugmented [accessed 2020-11-09]
19. Buchanan L, Yeatman H, Kelly B, Kariippanon K. A thematic content analysis of how marketers promote energy drinks on digital platforms to young Australians. Aust N Z J Public Health 2018 Dec;42(6):530-531. [doi: 10.1111/1753-6405.12840] [Medline: 30370962]
20. Telemedicine and COVID-19. GitHub. URL: https://github.com/cpollack736/telemedicine_covid19 [accessed 2021-06-09]

21. Office for Human Research Protections (OHRP). 2018 Requirements (2018 Common Rule). U.S. Department of Health & Human Services. 2017 Jan 19. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html [accessed 2021-05-19]
22. Ghebreyesus T. 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 2021-01-10]
23. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in outpatient care delivery and telemedicine during the COVID-19 pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391. [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]
24. Vidal-Alaball J, Fernandez-Luque L, Marin-Gomez FX, Ahmed W. A new tool for public health opinion to give insight into telemedicine: Twitter poll analysis. JMIR Form Res 2019 May 28;3(2):e13870 [FREE Full text] [doi: 10.2196/13870] [Medline: 31140442]
25. The COVID-19 Healthcare Coalition Telehealth Workgroup. Telehealth impact study: patient survey executive summary. The MITRE Corporation. Bedford, Massachusetts, USA; 2021 Apr 16. URL: https://c19hcc.org/static/catalog-resources/telehealth-patient-survey-analysis-c19hcc.pdf [accessed 2021-06-14]
26. Improving rural health and telehealth access. Federal Register - The Daily Journal of the United States Government. 2020 Aug 3. URL: https://www.federalregister.gov/documents/2020/08/06/2020-17364/improving-rural-health-and-telehealth-access [accessed 2021-06-11]
27. Shatzkes M, Kraus E, Rai K. It’s official: telehealth benefits have been expanded for Medicare beneficiaries. The National Law Review. 2020 Aug 10. URL: https://www.natlawreview.com/article/it-s-official-telehealth-benefits-have-been-expanded-medicare-beneficiaries [accessed 2021-05-20]
28. Latifi R, Doarn C. Perspective on COVID-19: finally, telemedicine at center stage. Telemed J E Health 2020 Sep;26(9):1106-1109. [doi: 10.1089/tmj.2020.0132] [Medline: 32408804]
29. Assistant Secretary for Public Affairs (ASPA). Telehealth: delivering care safely during COVID-19. U.S. Department of Health & Human Services. 2020 Jul 15. URL: https://www.hhs.gov/coronavirus/telehealth/index.html [accessed 2021-05-20]
30. 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]
31. 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]
32. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. 2019 Apr 10. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/ [accessed 2021-05-19]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
GUI: graphical user interface
IRR: incidence risk ratio
NCI: National Cancer Institute
NIH: National Institutes of Health
NLM: National Library of Medicine
VADER: Valence Aware Dictionary and sEntiment Reasoner

Edited by C Basch; submitted 09.03.21; peer-reviewed by J Burzyńska, R Gore; comments to author 17.04.21; revised version received 31.05.21; accepted 01.06.21; published 21.06.21.

Please cite as:
Pollack CC, Gilbert-Diamond D, Alford-Teaster JA, Onega T
Language and Sentiment Regarding Telemedicine and COVID-19 on Twitter: Longitudinal Infodemiology Study
J Med Internet Res 2021;23(6):e28648
URL: https://www.jmir.org/2021/6/e28648
doi:10.2196/28648
PMID:34086591

Original Paper

Engagement With COVID-19 Public Health Measures in the United States: A Cross-sectional Social Media Analysis from June to November 2020

Daisy Massey[1], BA; Chenxi Huang[1], PhD; Yuan Lu[1,2], ScD; Alina Cohen[3], MS, MBA; Yahel Oren[3], MS; Tali Moed[3], BS; Pini Matzner[3], PhD; Shiwani Mahajan[1,2], MBBS, MHS; César Caraballo[1,2], MD; Navin Kumar[4], MPhil; Yuchen Xue[5], MA; Qinglan Ding[6], PhD, MBBS; Rachel Dreyer[7], PhD; Brita Roy[8,9], MD, MPH, MHS; Harlan Krumholz[1,2,10], MD, SM

[1]Center for Outcomes Research and Evaluation, Yale New Haven Hospital, New Haven, CT, United States
[2]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, United States
[3]Signals Analytics, New York, NY, United States
[4]Department of Sociology, Yale University, New Haven, CT, United States
[5]Foundation for a Smoke-Free World, New York, NY, United States
[6]College of Health and Human Sciences, Purdue University, West Lafayette, IN, United States
[7]Department of Emergency Medicine, Yale School of Medicine, New Haven, CT, United States
[8]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, United States
[9]Department of Medicine, Yale School of Medicine, New Haven, CT, United States
[10]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, United States

Corresponding Author:
Harlan Krumholz, MD, SM
Section of Cardiovascular Medicine
Department of Internal Medicine
Yale School of Medicine
1 Church St
Suite 200
New Haven, CT
United States
Phone: 1 203 764 5885
Email: harlan.krumholz@yale.edu

Abstract

Background: COVID-19 has continued to spread in the United States and globally. Closely monitoring public engagement and perceptions of COVID-19 and preventive measures using social media data could provide important information for understanding the progress of current interventions and planning future programs.

Objective: The aim of this study is to measure the public's behaviors and perceptions regarding COVID-19 and its effects on daily life during 5 months of the pandemic.

Methods: Natural language processing (NLP) algorithms were used to identify COVID-19–related and unrelated topics in over 300 million online data sources from June 15 to November 15, 2020. Posts in the sample were geotagged by NetBase, a third-party data provider, and sensitivity and positive predictive value were both calculated to validate the classification of posts. Each post may have included discussion of multiple topics. The prevalence of discussion regarding these topics was measured over this time period and compared to daily case rates in the United States.

Results: The final sample size included 9,065,733 posts, 70% of which were sourced from the United States. In October and November, discussion including mentions of COVID-19 and related health behaviors did not increase as it had from June to September, despite an increase in COVID-19 daily cases in the United States beginning in October. Additionally, discussion was more focused on daily life topics (n=6,210,255, 69%), compared with COVID-19 in general (n=3,390,139, 37%) and COVID-19 public health measures (n=1,836,200, 20%).

Conclusions: There was a decline in COVID-19–related social media discussion sourced mainly from the United States, even as COVID-19 cases in the United States increased to the highest rate since the beginning of the pandemic. Targeted public health messaging may be needed to ensure engagement in public health prevention measures as global vaccination efforts continue.

(J Med Internet Res 2021;23(6):e26655) doi:10.2196/26655

KEYWORDS

COVID-19; public perception; social media; infodemiology; infoveillance; infodemic; social media research; social listening; social media analysis; natural language processing; Reddit data; Facebook data; COVID-19 public health measures; public health; surveillance; engagement; United States; cross-sectional; Reddit; Facebook; behavior; perception; NLP

Introduction

As COVID-19 continues its spread in the United States, a key to controlling the spread while vaccination efforts continue is to enlist the public in risk-mitigation behaviors [1,2]. Studying the public's social media posts regarding COVID-19 public health measures may provide information about targets of interventions, progress toward behavior goals, and the risk of future outbreaks [3-9]. Although real-time reports on pandemic-related tests and mortality are widely available, there are fewer opportunities to gain near real-time insight into behaviors and beliefs about the pandemic.

Social media, which people are using now more than ever to communicate, has served as a useful data source for providing rapid insight into the public's behaviors and beliefs during the pandemic [10-13]. Studies have noted a high prevalence of COVID-19–related discussion—including such topics as hygiene, shortages, and the spread of misinformation—and an increase in COVID-19–related discussion as COVID-19 cases increase [5,14,15]. However, existing findings are based on evidence during only the beginning of the outbreak, from December 2019 to April 2020, and the range of topics and keywords explored is also limited [7,14-19]. Additionally, studies analyzing COVID-19 behaviors and beliefs on social media have primarily used Twitter as their source, which has several limitations [14-16,19]. Most notably, highly rated retweets are more likely to come from spam and bot accounts, which are also actively posting about COVID-19, and can obscure the targeting of signals from human discussions [20-22]. Further, previous studies each focused on a particular aspect of the pandemic, such as misinformation relating to the pandemic, without comparing the volume of discussion related to multiple aspects to determine the public's relative focus on particular pandemic-related issues and behaviors. Therefore, there is a need to assess how the public's current reaction to the pandemic has changed since the early stages, by examining broad online discussion from more diverse sources.

Accordingly, we measured the prevalence of online discussion that included topics in the categories of daily life, which may or may not be related to COVID-19, and COVID-19–related public health, from June through November 2020. We also assessed the correlation between prevalence of discussion topics and US COVID-19 new daily case rates (incidence). In measuring these trends in social media data and the COVID-19 incidence rate in the United States, we sought to elucidate the US public's engagement with COVID-19–related public health measures, which are crucial to addressing the current pandemic.

Methods

Data Sources

The data sample consisted of unstructured, English-language posts from forums, such as Reddit, Facebook public pages, and 4Chan, and comments from news sites (Table S1 in Multimedia Appendix 1) [23]. We defined forums as thread lists or topic-specific pages, and excluded social media sites including Twitter, YouTube, Instagram, and LinkedIn [24]. Signals Analytics, an advanced analytics consulting firm that conducted the analysis, accessed these data sources through a third-party data vendor, NetBase [25,26]. These social media posts were geotagged by NetBase both directly, by using geolocation data from posts, and indirectly, by using author profiles and unique domain codes (such as .uk). All data were deidentified by NetBase before being transferred to Signals Analytics.

In addition to the social data, the study included US COVID-19 case data from the COVID-19 Dashboard by the Center for Systems Science and Engineering at Johns Hopkins University [27]. These data were updated daily using a public application programming interface (API) and included total number of deaths, new daily deaths, total active cases, and daily new cases [28].

No personal identifying information (eg, usernames, emails, or IP addresses) was shared as part of the analysis or reporting process. This study was exempted from Institutional Review Board review by Yale University as it did not engage in research involving human subjects.

Approach

To determine trends in social media discussion during the COVID-19 pandemic, we collected data posts from all internet sources and applied natural language processing (NLP) algorithms to identify and classify mentions of COVID-19, COVID-19–related public health measures, and daily life topics (Table S2 in Multimedia Appendix 1).

NetBase ran a daily query that we designed based on our project scope on over 300 million online data sources from June 15 to November 15, 2020 (Methods 1 in Multimedia Appendix 1). There were several steps to narrow the sample retrieved from the query to include only posts relevant to our research question (Figure S1 in Multimedia Appendix 1). First, NLP algorithms were run to remove advertisements and pornography-related sites and posts (Methods 2 in Multimedia Appendix 1). Next, a taxonomy of topics was applied (Methods 3 in Multimedia Appendix 1). The posts that did not include discussion of topics from the taxonomy were deleted. Finally, all news articles and blog posts were deleted from the sample, so that the only remaining data posts were from social outlets (forums and comments on news sites).

The taxonomy was comprised of two categories, COVID-19–related public health measures and daily life behaviors, each of which included multiple topics (Methods 4 in Multimedia Appendix 1). COVID-19 mentions was also an individual topic in the taxonomy, independent of either category. Any post that directly mentioned COVID-19 by name or synonym, including slang such as "Miss Rona," was classified as including a COVID-19 mention (Table S2 in Multimedia Appendix 1). Taxonomy categories and topics were not exclusive, so that a post was classified as belonging to each taxonomy topic and category that it contained mention of (Table S2 in Multimedia Appendix 1).

Once all posts were classified according to the topics in the taxonomy, we measured trends in these topics over time by tracking the total number of posts that included mentions of each taxonomy topic and category. Classifications of topics and categories were not mutually exclusive, so the same post was able to be classified into multiple topics across any category. Trends were visualized by taxonomy category, COVID-19 mentions, and by the most commonly mentioned taxonomy topics. These trends were visualized with the COVID-19 incidence rate in the United States. We chose to correlate the trends in taxonomy topics with trends in the COVID-19 incidence rate rather than the COVID-19 death rate based on previous literature, which found a correlation between trends in online social chatter and COVID-19 incidence [3,5].

This approach allowed us to identify changes in both topics that prior research in the early stage of the outbreak had shown to be prevalent in COVID-19 discussion, and topics from daily life and COVID-19 literature reviews that were not previously known to be found in COVID-19 discussion, but that may have become apparent as COVID-19 cases or current events changed [15,16,29-33]. Additionally, our approach removed redundant posts, limiting the effect of bots and reposts (Methods 3 in Multimedia Appendix 1). The taxonomy classification was validated by calculating positive predictive value and sensitivity (Methods 5 in Multimedia Appendix 1). We also validated the methodology by applying it to US-specific current events and found that the approach revealed an increase in online social discussion when the given current event topic was most relevant (Figure S2 in Multimedia Appendix 1). This methodology was shown to reveal insights into outbreak characterization and event prediction for the e-cigarette or vaping use–associated lung injury outbreak [34].

Results

The final data sample consisted of 9,065,733 online social posts that mentioned at least one of the topics in our taxonomy from June 15 to November 15, 2020 (Table 1). The majority (87%) of posts in our sample came from sources that were categorized as forums, including Reddit, Facebook, and 4Chan (Table 2; Table S1 in Multimedia Appendix 1) [23]. The minority of posts (13%) in our sample were derived from comment sections on news sites, including The Hill, a media source focused on politics and business, and Breitbart, a right-leaning media source (Table 2; Table S1 in Multimedia Appendix 1) [35,36]. Most posts in the sample were not able to be directly geotagged due to sources' data privacy measures and restrictions. A minority were geotagged as from the United States, with the remaining geotagged as from a country other than the United States (Table S3 in Multimedia Appendix 1). Using indirect geotagging provided by NetBase, it was estimated that about 70% of all initial posts collected by the search query were from the United States. In an independent data sample of 100 posts classified by manual review, the algorithm had a positive predictive value of over 80%, which was calculated as the number of posts correctly classified by the taxonomy using NLP algorithms divided by the number of all posts classified by the taxonomy. This was a higher accuracy measure than is found in comparable social media research [30]. Sensitivity was calculated as the number of correct classifications of a topic using the NLP algorithms divided by the total number of posts for the topic identified by manual screening, and we found that our taxonomy approach led to an average classification rate of 92% sensitivity.

Within the data sample, 6,210,255 (69%) posts were classified as including discussion of daily life topics, while 3,390,139 (37%) contained mentions of COVID-19, and 1,836,200 (20%) posts were classified as including discussion of COVID-19–related public health topics (Table 1). The most prevalent topics among the daily life posts were sex life (n=887,457, 14%), food (n=838,513, 14%), and financial concerns (n=710,757, 11%). The most prevalent topic in COVID-19–related public health behaviors posts was wearing face masks (n=1,120,344, 61%), followed by lockdowns (n=457,705, 25%), and social distancing (n=242,105, 13%).

Online social posts including COVID-19 mentions and discussion of COVID-19–related public health behaviors increased in June 2020, as COVID-19 cases also increased, but remained stagnant as cases began to increase in October (Figure 1). Discussion about wearing face masks was most prevalent in mid-July, during the summer wave (mid-June to early September) of COVID-19 cases, and remained at pre-June levels in October and November, with the exception of a sharp increase on October 2, 2020 (Figure 2).

Table 1. Number of posts by taxonomy topic from June 15 to November 15, 2020 (N=9,065,733)[a].

Relevant taxonomy categories (percent classified within all posts) and topics	Number of posts with mentions (percent classified within category)
COVID-19–related public health topics (20)	1,836,200
Wearing face mask	1,120,344 (61)
Lockdown	457,705 (25)
Social distancing	242,105 (13)
Quarantine	94,301 (5)
Testing	87,712 (5)
Excessive handwashing	64,679 (4)
Contact tracing	31,775 (2)
Reopening	16,681 (1)
Screening	14,569 (1)
Wearing gloves	11,531 (1)
Disinfection	11,076 (1)
Wearing face shield	10,104 (1)
Daily life taxonomy topics (69)	6,210,255
Sex life	887,457 (14)
Food	838,513 (14)
Financial	710,757 (11)
Travel	651,426 (10)
Smoking/vaping	476,468 (8)
Mass gatherings	451,815 (7)
Virtual communication	414,549 (7)
Alcohol consumption	398,229 (6)
Religion	285,538 (5)
New skills/hobbies acquisition/DIY	280,155 (5)
Drug use	257,819 (4)
News/media consumption	257,415 (4)
Reading	246,074 (4)
Physical activity	205,116 (3)
Work from home	198,057 (3)
Socializing in person	177,522 (3)
Stockpiling	171,421 (3)
Relaxation techniques	164,262 (3)
Excess sleep	127,623 (2)
Pets	109,510 (2)
Postponing plans	98,626 (2)
Childcare	97,735 (2)
Public transportation	94,414 (2)
Reduced sleep quality	88,196 (1)
Home school	80,153 (1)
Non–COVID-19 hospital visits	77,278 (1)
Doctor well visit	72,235 (1)
Funerals	45,394 (1)

Relevant taxonomy categories (percent classified within all posts) and topics	Number of posts with mentions (percent classified within category)
Family-centered time	28,106 (0)
Outdoor culture	21,546 (0)
Births	17,283 (0)
Telehealth	11,574 (0)
Smokeless tobacco consumption	2136 (0)
COVID-19 mentions (37)	3,390,139

[a] Percentages do not sum to 100 because each post may have included discussion of multiple topics, including topics in different categories.

Table 2. Number of posts by source type from June 15 to November 15, 2020.

Category of COVID-19 discussion topic	Posts with mentions of COVID-19–related public health behavior (N=1,836,200), n (%)	Posts with mentions of daily life (N=6,210,255), n (%)	Posts with mentions of COVID-19 (N=3,390,139), n (%)	Total data sample (N=9,065,733), n (%)
Forums	1,494,401 (81)	5,714,446 (92)	2,749,451 (81)	7,928,599 (87)
Comments	341,799 (19)	495,809 (8)	640,688 (19)	1,137,134 (13)

Figure 1. Online social discussion categories versus US daily new COVID-19 cases (June 15 to November 15, 2020).

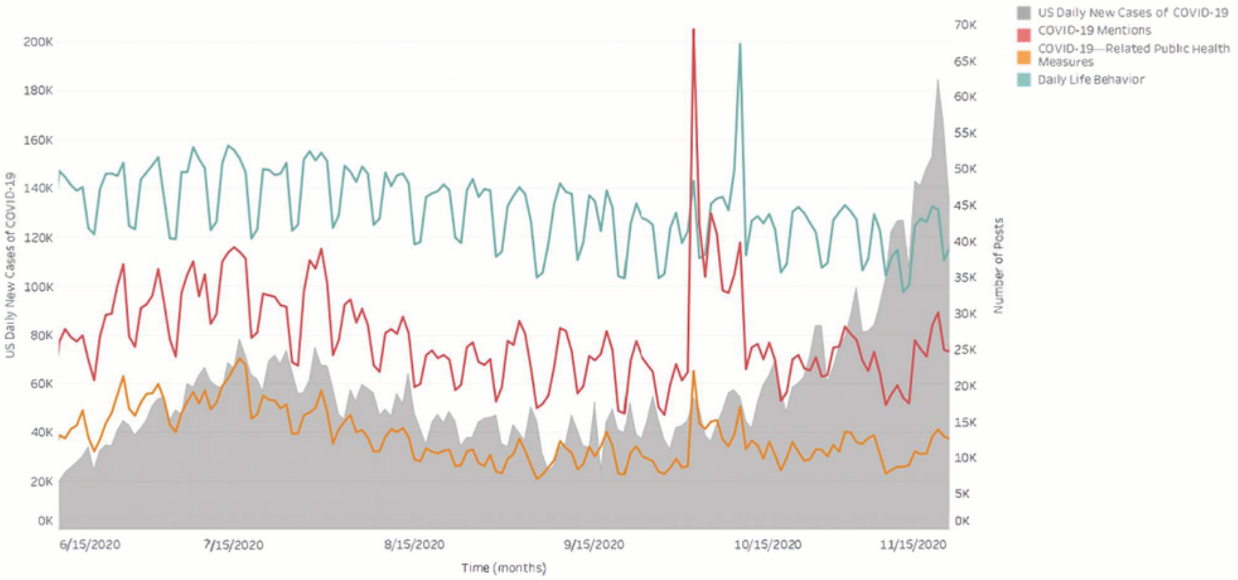
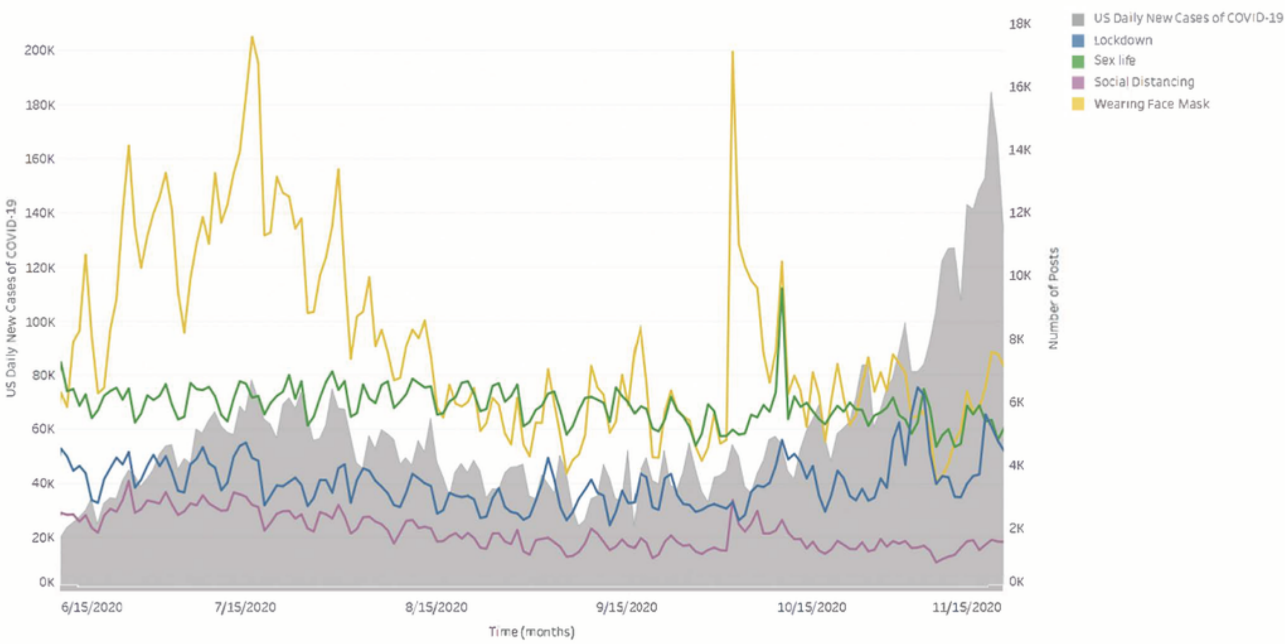


Figure 2. Public health measures online social discussion versus US daily new COVID-19 cases (June 15 to November 15, 2020).



Discussion

Principal Findings

Our study had several important findings. From June to November 2020, predominantly US-based online social chatter was more focused on daily life than it was on public health behaviors relating to COVID-19. In addition, although discussion relating to COVID-19 and related public health behaviors appeared to increase with rising US cases in the summer wave (early June to early September), the volume of COVID-19–related discussion was lower in the wave that began in the fall (mid-October), despite the fact that, during the fall wave, COVID-19 cases increased to their highest rates since the pandemic began [37]. In particular, discussion of wearing face masks, the most prevalent of any COVID-19 public health behavior we studied, declined in mid-July despite the pandemic continuing and evidence that wearing face masks has not been universally adopted in the United States, and increased only minimally once cases began to increase again in early October [38,39]. One exception to this finding was the brief but stark increase in COVID-19–related discussion on October 2, 2020, which coincided with the announcement that President Donald Trump had contracted COVID-19 [40]. Our finding that daily life topics were more prevalent in social media chatter than COVID-19–related public health behaviors and mentions of COVID-19 is not immediately surprising given the differences in scope. Nevertheless, we applied consistent methods over time, and the decrease of COVID-19–related discussion in the context of the fall rise in COVID-19 cases differs from the pattern we visualized in the summer wave.

Our study expanded upon previous COVID-19–related social media analyses in that our sources used forums and comments on news sites instead of Twitter and our study was conducted in later phases of the pandemic. Our study sources included forums and comments on news sites, which we believe was an advantage for a few reasons. First, forums are unique to other forms of social media in that they tend to include more text, with greater character allowances and less frequent use of hashtags. This allows the NLP algorithms to be more accurately applied, because forum users include more context to which inclusion and exclusion criteria can be applied. Reddit has also been found to include more discussion than links to external sources, again providing more context to analyze [41]. Second, forums, such as Reddit and public Facebook pages, and comments on news sites, are already focused on specific topics and therefore have more in-depth discussions on the same topic, as opposed to other social media sites, which more often share updates from individual users or links to other sites. The added context from in-depth discussions also allows for more accurate NLP classification. Third, as discussed earlier, retweets driven by spam and bot accounts on Twitter can obscure the targeting of signals from human discussion [20-22].

Due to these differences in study design and time period, our findings may not be consistent with those of previous studies from the first wave of the COVID-19 pandemic. However, future research may investigate whether the cause of the different findings is a significant difference between the type of social chatter found on forums and that found on Twitter and other social media platforms, or whether the different findings are due to a temporal trend of a decreased focus on COVID-19. Although we found that online social chatter was more focused on daily life than it was on COVID-19 public health behaviors, previous research found the opposite. For instance, one study from March 2020 that used data from Twitter found that social media discussion about COVID-19–related health topics was more common than discussion about daily life topics such as socializing, the economy, or politics [42]. Earlier research also found that COVID-19–related public health measures were discussed not only more often than social topics, but also more often than other COVID-19–related topics [7,15]. Thus, our finding that online social chatter from June to November was more focused on daily life than it was on COVID-19 public health behaviors may indicate that the public's focus on COVID-19 preventative health behaviors had decreased since previous studies were conducted in March and April, or our results may have differed from these earlier studies because our study used different data sources and excluded Twitter. There have been related studies that have analyzed social media data on Reddit—a major source of data in our analysis—during the pandemic; however, none of these studies addressed our research question directly, which was how levels of COVID-19–related public health discussion compared to levels of daily life and COVID-19–related discussion over time. We noted three studies conducted during the time period from January to May 2020 discovered and measured common COVID-19–related topics among online Reddit posts without determining the relative prevalence of COVID-19–related public health discussion to daily life discussion [43-45]. One additional study found that, from February to May 2020, there was a positive correlation between COVID-19–related news coverage and COVID-19–related discussion on the r/Coronavirus subreddit, but that the COVID-19–related discussion declined after sustained media coverage, showing that public attention saturates [46].

Although our results cannot be compared to previous studies to show that public perception changed from the spring wave to the summer and fall waves, there is precedent for the interpretation that the public's focus on COVID-19 public health measures waned during the fall months. As public health experts warned against relaxing preventive behaviors as pandemic fatigue grew, activity and traffic data indicated that people may have stopped adhering to public health recommendations to stay home and avoid close contact with people outside their household [47-50]. The decline of chatter regarding wearing face masks, and the relative low rates of discussions on other COVID-19–related public health behaviors, may reflect that social media engagement with these issues decreased as the pandemic progressed, and remained low among the US population as the pandemic continued to confront a high COVID-19 daily case rate.

Our study has several limitations. First, although our third-party data provider, NetBase, reported that about 70% of posts were from the United States based on indirect geotagging methods, we do not know the location for most posts according to our direct geotagging methods, which were only able to tag about 20% of posts (Table S3 in Multimedia Appendix 1). As a result, we cannot make international comparisons, but our data set is more representative of the United States than of any other country. Second, the number of posts included in our data set was much lower than previous studies, likely due to the types of data sources used, which excluded social media sites such as Twitter in order to exclude noise that might have obscured signals in data, and our methodology, which included removing posts not relevant to our more refined taxonomy. We used a stringent exclusion criterion with a list of prespecified keywords that may also have led to a smaller sample size, but our approach aimed to create a sample with high accuracy levels. Third, we were not able to include sentiment analysis or other content analysis in our study, which is an area for further exploration. Finally, there is no demographic information available from the data posts directly due to privacy considerations and data use agreements. Thus, we cannot determine whether our data sample contains biases due to the demographics of the people who posted. For instance, Reddit, which was the most common forum source for our data sample, has been found to be used by a younger, male audience [51,52].

Conclusion

In this study of predominantly US-based COVID-19 social media data from June to November 2020, we observed that COVID-19 and relevant public health measures were discussed less than daily life behaviors on social media, and that discussion on wearing face masks decreased throughout the summer and into the fall, while cases increased. These discussion rates may reveal a need for increased public health messaging as the pandemic continues.

Acknowledgments

This work was supported by the project Insights about the COVID Pandemic Using Public Data IRES PD: 20-005872, with funding from the Foundation for a Smoke-Free World.

Authors' Contributions

AC, TM, PM, and YO from Signals Analytics had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. DM from Yale School of Medicine takes full responsibility for the data interpretation and writing. CH from Yale School of Medicine contributed to analyzing the data. YL, SM, CC, NK, YX, QD, RD, BR, HK contributed to editing the manuscript.

Conflicts of Interest

YL is supported by the National Heart, Lung, and Blood Institute (K12HL138037) and the Yale Center for Implementation Science. RD is supported by an American Heart Association Transformational Project Award (#19TPA34830013) and a Canadian Institutes of Health Research Project Grant (RN356054–401229). In the past three years, HK received expenses and/or personal fees from UnitedHealth, IBM Watson Health, Element Science, Aetna, Facebook, the Siegfried and Jensen Law Firm, Arnold and Porter Law Firm, Martin/Baughman Law Firm, F-Prime, and the National Center for Cardiovascular Diseases in Beijing. He is an owner of Refactor Health and HugoHealth, and had grants and/or contracts from the Centers for Medicare & Medicaid Services, Medtronic, the U.S. Food and Drug Administration, Johnson & Johnson, and the Shenzhen Center for Health Information. The remaining authors have no disclosures to report.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 172 KB - jmir_v23i6e26655_app1.docx]

References

1. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 2020 Mar 21;395(10228):931-934 [FREE Full text] [doi: 10.1016/S0140-6736(20)30567-5] [Medline: 32164834]
2. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA 2020 May 19;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
3. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
4. Husnayain A, Shim E, Fuad A, Su ECY. Assessing the community risk perception toward COVID-19 outbreak in South Korea: evidence from Google and NAVER relative search volume. MedRxiv. Preprint published online on April 29, 2020. [doi: 10.1101/2020.04.23.20077552]
5. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
6. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593. [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
7. Wang X, Zou C, Xie Z, Li D. Public Opinions towards COVID-19 in California and New York on Twitter. medRxiv. Preprint published online on July 14, 2020 [FREE Full text] [doi: 10.1101/2020.07.12.20151936] [Medline: 32699856]
8. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
9. Malecki K, Keating JA, Safdar N. Crisis Communication and Public Perception of COVID-19 Risk in the Era of Social Media. Clin Infect Dis 2021 Feb 16;72(4):697-702 [FREE Full text] [doi: 10.1093/cid/ciaa758] [Medline: 32544242]
10. Koeze E, Popper N. The Virus Changed the Way We Internet. The New York Times. 2020 Apr 07. URL: https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html [accessed 2021-06-15]
11. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital Health Strategies to Fight COVID-19 Worldwide: Challenges, Recommendations, and a Call for Papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
12. Li S, Feng B, Liao W, Pan W. Internet Use, Risk Awareness, and Demographic Characteristics Associated With Engagement in Preventive Behaviors and Testing: Cross-Sectional Survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]
13. Brady WJ, Crockett MJ, Van Bavel JJ. The MAD Model of Moral Contagion: The Role of Motivation, Attention, and Design in the Spread of Moralized Content Online. Perspect Psychol Sci 2020 Jul;15(4):978-1010. [doi: 10.1177/1745691620917336] [Medline: 32511060]
14. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]

15. Singh L, Bansal S, Bode L, Budak C, Chi G, Kawintiranon K, et al. A first look at COVID-19 information and misinformation sharing on Twitter. ArXiv. Preprint posted online on March 31, 2020. [Medline: 32550244]
16. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
17. Massaad E, Cherfan P. Social Media Data Analytics on Telehealth During the COVID-19 Pandemic. Cureus 2020 Apr 26;12(4):e7838 [FREE Full text] [doi: 10.7759/cureus.7838] [Medline: 32467813]
18. Mackey TK, Li J, Purushothaman V, Nali M, Shah N, Bardier C, et al. Big Data, Natural Language Processing, and Deep Learning to Detect and Characterize Illicit COVID-19 Product Sales: Infoveillance Study on Twitter and Instagram. JMIR Public Health Surveill 2020 Aug 25;6(3):e20794 [FREE Full text] [doi: 10.2196/20794] [Medline: 32750006]
19. Mackey T, Purushothaman V, Li J, Shah N, Nali M, Bardier C, et al. Machine Learning to Detect Self-Reporting of Symptoms, Testing Access, and Recovery Associated With COVID-19 on Twitter: Retrospective Big Data Infoveillance Study. JMIR Public Health Surveill 2020 Jun 08;6(2):e19509 [FREE Full text] [doi: 10.2196/19509] [Medline: 32490846]
20. Chu Z, Gianvecchio S, Wang H, Jajodia S. Who is tweeting on Twitter: human, bot, or cyborg? In: Proceedings of the 26th Annual Computer Security Applications Conference.: Association for Computing Machinery; 2010 Presented at: 26th Annual Computer Security Applications Conference; December 6-10, 2010; Austin, TX URL: https://www.eecis.udel.edu/~hnw/paper/acsac10.pdf
21. Tsou M, Zhang H, Jung CT. Identifying data noises, user biases, and system errors in geo-tagged Twitter messages (Tweets). ArXiv. Preprint posted online on December 6, 2017 [FREE Full text]
22. Ferrara E. What Types of COVID-19 Conspiracies are Populated by Twitter Bots? ArXiv. Preprint posted online on April 20, 2020 [FREE Full text]
23. Bernstein M, Monroy-Hernández A, Harry D, André P, Panovich K, Vargas G. 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2873/4398
24. Weichselbraun A, Brasoveanu AMP, Waldvogel R, Odoni F. Harvest--An Open Source Toolkit for Extracting Posts and Post Metadata from Web Forums. ArXiv. Preprint posted online on February 3, 2021 [FREE Full text]
25. Signals Analytics. URL: https://www.signals-analytics.com [accessed 2020-12-11]
26. NetBase Quid. URL: https://netbasequid.com [accessed 2020-12-11]
27. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
28. Axisbits. COVID-19 Statistics API Documentation. Rapid API. URL: https://rapidapi.com/axisbits-axisbits-default/api/covid-19-statistics/details [accessed 2021-06-15]
29. Han X, Wang J, Zhang M, Wang X. Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. Int J Environ Res Public Health 2020 Apr 17;17(8):1 [FREE Full text] [doi: 10.3390/ijerph17082788] [Medline: 32316647]
30. 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]
31. Nierenberg A, Pasick A. Schools Reopening: The State of Play for K-12. The New York Times. 2020 Aug 17. URL: https://www.nytimes.com/2020/08/17/us/k-12-schools-reopening.html [accessed 2021-06-15]
32. Goodnough A, Sheikh K. CDC Weighs Advising Everyone to Wear a Mask. The New York Times. 2020 Mar 31. URL: https://www.nytimes.com/2020/03/31/health/cdc-masks-coronavirus.html [accessed 2021-06-15]
33. Sheikh K. You're Getting Used to Masks. Will You Wear a Face Shield? The New York Times. 2020 May 24. URL: https://www.nytimes.com/2020/05/24/health/coronavirus-face-shields.html [accessed 2021-06-15]
34. Matzner P. Using Advanced Analytics for the Early Detection of Pandemics and Outbreaks. Signals Analytics. URL: https://info.signals-analytics.com/research-paper-early-detection-pandemics [accessed 2021-06-15]
35. The Hill Media Bias | AllSides. AllSides. URL: https://www.allsides.com/news-source/hill-media-bias [accessed 2021-06-15]
36. Ribeiro F, Henrique L, Benevenuto F, Chakraborty A, Kulshrestha J, Babaei M. Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale. In: Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018). 2018 Presented at: Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018); June 25-28, 2018; Palo Alto, CA URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17878/17020
37. Wan W, Dupree J. U.S. hits highest daily number of coronavirus cases since pandemic began. Washington Post. 2020 Oct 23. URL: https://www.washingtonpost.com/health/2020/10/23/covid-us-spike-cases/ [accessed 2021-06-15]
38. Katz J, Sanger-Katz M, Quealy K. A Detailed Map of Who Is Wearing Masks in the US. The New York Times. 2020 Jul 17. URL: https://www.nytimes.com/interactive/2020/07/17/upshot/coronavirus-face-mask-map.html [accessed 2021-06-15]
39. Brenan M. Americans' Face Mask Usage Varies Greatly by Demographics. Gallup. 2020 Jul 13. URL: https://news.gallup.com/poll/315590/americans-face-mask-usage-varies-greatly-demographics.aspx [accessed 2021-06-15]

40. Baker P, Haberman M. Trump Tests Positive for the Coronavirus. The New York Times. 2020 Oct 02. URL: https://www.nytimes.com/2020/10/02/us/politics/trump-covid.html [accessed 2021-06-15]
41. Singer P, Flöck F, Meinhart C, Zeitfogel E, Strohmaier M. Evolution of reddit: from the front page of the internet to a self-referential community? ArXiv. Preprint posted online on February 6, 2014 [FREE Full text]
42. Molla R. How coronavirus took over social media. Vox. 2020 Mar 12. URL: https://www.vox.com/recode/2020/3/12/21175570/coronavirus-covid-19-social-media-twitter-facebook-google [accessed 2021-06-15]
43. Stokes D, Andy A, Guntuku SC, Ungar LH, Merchant RM. Public Priorities and Concerns Regarding COVID-19 in an Online Discussion Forum: Longitudinal Topic Modeling. J Gen Intern Med 2020 Jul;35(7):2244-2247 [FREE Full text] [doi: 10.1007/s11606-020-05889-w] [Medline: 32399912]
44. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
45. Chipidza W, Akbaripourdibazar E, Gwanzura T, Gatto NM. Topic Analysis of Traditional and Social Media News Coverage of the Early COVID-19 Pandemic and Implications for Public Health Communication. Disaster Med Public Health Prep 2021 Mar 03:1-8 [FREE Full text] [doi: 10.1017/dmp.2021.65] [Medline: 33653437]
46. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]
47. Miller SG, Weaver J. Fauci Says U.S. Won't Get Back to Normal Until Late 2021. NBC News. 2020 Sep 11. URL: https://www.nbcnews.com/health/health-news/fauci-says-us-won-t-get-back-normal-until-late-n1239882 [accessed 2021-06-15]
48. Ghader S, Zhao J, Lee M, Zhou W, Zhao G, Zhang L. Observed mobility behavior data reveal social distancing inertia. ArXiv. Preprint posted online on April 30, 2020 [FREE Full text]
49. Traffic Monitoring Count Data: Volume and Classification Information. CT.gov. URL: https://portal.ct.gov/DOT/PP_SysInfo/Traffic-Monitoring [accessed 2021-06-15]
50. Schuman R. INRIX U.S. National Traffic Volume Synopsis Issue #15 (June 20 – June 26, 2020). INRIX. 2020 Jun 29. URL: https://inrix.com/blog/2020/06/covid19-us-traffic-volume-synopsis-15/ [accessed 2021-06-15]
51. Inferring gender of Reddit users. Vasilev E. 2018. URL: https://kola.opus.hbz-nrw.de/opus45-kola/frontdoor/deliver/index/docId/1619/file/Master_thesis_Vasilev.pdf [accessed 2021-06-15]
52. Finlay SC. Age and Gender in Reddit Commenting and Success. Journal of Information Science Theory and Practice 2014 Sep 30;2(3):18-28. [doi: 10.1633/jistap.2014.2.3.2]

Abbreviations

API: application programming interface
NLP: natural language processing

Edited by C Basch; submitted 20.12.20; peer-reviewed by G Aguayo, M Nali; comments to author 01.02.21; revised version received 05.03.21; accepted 16.04.21; published 21.06.21.

Please cite as:
Massey D, Huang C, Lu Y, Cohen A, Oren Y, Moed T, Matzner P, Mahajan S, Caraballo C, Kumar N, Xue Y, Ding Q, Dreyer R, Roy B, Krumholz H
Engagement With COVID-19 Public Health Measures in the United States: A Cross-sectional Social Media Analysis from June to November 2020
J Med Internet Res 2021;23(6):e26655
URL: https://www.jmir.org/2021/6/e26655
doi:10.2196/26655
PMID:34086593

Original Paper

COVID-19 Vaccine–Related Discussion on Twitter: Topic Modeling and Sentiment Analysis

Joanne Chen Lyu[1], PhD; Eileen Le Han[2], PhD; Garving K Luli[3], PhD

[1]Center for Tobacco Control Research and Education, University of California, San Francisco, San Francisco, CA, United States
[2]School of Information, University of Michigan, Ann Arbor, MI, United States
[3]Department of Mathematics, University of California, Davis, Davis, CA, United States

Corresponding Author:
Joanne Chen Lyu, PhD
Center for Tobacco Control Research and Education
University of California, San Francisco
530 Parnassus Avenue
San Francisco, CA, 94143-1390
United States
Phone: 1 415 502 4181
Email: chenjoanne.lyu@ucsf.edu

Related Article:

Comment in: https://www.jmir.org/2022/2/e31978

Abstract

Background: Vaccination is a cornerstone of the prevention of communicable infectious diseases; however, vaccines have traditionally met with public fear and hesitancy, and COVID-19 vaccines are no exception. Social media use has been demonstrated to play a role in the low acceptance of vaccines.

Objective: The aim of this study is to identify the topics and sentiments in the public COVID-19 vaccine–related discussion on social media and discern the salient changes in topics and sentiments over time to better understand the public perceptions, concerns, and emotions that may influence the achievement of herd immunity goals.

Methods: Tweets were downloaded from a large-scale COVID-19 Twitter chatter data set from March 11, 2020, the day the World Health Organization declared COVID-19 a pandemic, to January 31, 2021. We used R software to clean the tweets and retain tweets that contained the keywords *vaccination*, *vaccinations*, *vaccine*, *vaccines*, *immunization*, *vaccinate*, and *vaccinated*. The final data set included in the analysis consisted of 1,499,421 unique tweets from 583,499 different users. We used R to perform latent Dirichlet allocation for topic modeling as well as sentiment and emotion analysis using the National Research Council of Canada Emotion Lexicon.

Results: Topic modeling of tweets related to COVID-19 vaccines yielded 16 topics, which were grouped into 5 overarching themes. Opinions about vaccination (227,840/1,499,421 tweets, 15.2%) was the most tweeted topic and remained a highly discussed topic during the majority of the period of our examination. Vaccine progress around the world became the most discussed topic around August 11, 2020, when Russia approved the world's first COVID-19 vaccine. With the advancement of vaccine administration, the topic of instruction on getting vaccines gradually became more salient and became the most discussed topic after the first week of January 2021. Weekly mean sentiment scores showed that despite fluctuations, the sentiment was increasingly positive in general. Emotion analysis further showed that trust was the most predominant emotion, followed by anticipation, fear, sadness, etc. The trust emotion reached its peak on November 9, 2020, when Pfizer announced that its vaccine is 90% effective.

Conclusions: Public COVID-19 vaccine–related discussion on Twitter was largely driven by major events about COVID-19 vaccines and mirrored the active news topics in mainstream media. The discussion also demonstrated a global perspective. The increasingly positive sentiment around COVID-19 vaccines and the dominant emotion of trust shown in the social media discussion may imply higher acceptance of COVID-19 vaccines compared with previous vaccines.

(J Med Internet Res 2021;23(6):e24435) doi:10.2196/24435

KEYWORDS

COVID-19; vaccine; vaccination; Twitter; infodemiology; infoveillance; topic; sentiment; opinion; discussion; communication; social media; perception; concern; emotion

Introduction

As the COVID-19 pandemic spread globally, COVID-19 vaccine-related issues received increasing public attention. Multiple research teams in major pharmaceutical companies and research institutions across various nations have been developing vaccines [1,2]. Although vaccination is a cornerstone of the prevention of communicable infectious diseases [3], the practice has traditionally faced public fears, hesitancy, and even opposition [4,5]. During the COVID-19 pandemic, it is estimated that 55% to 85% of the population, depending on the country and the infection rate, needs to receive the COVID-19 vaccine to provide herd immunity [6,7]. However, a survey about COVID-19 vaccine intentions in September 2020 suggested that 21% of the public in the United States would definitely get vaccinated and 24% would definitely not get vaccinated [8]. One factor leading to the low acceptance of vaccines is poor health literacy, which is significantly influenced by social media use [9]. Therefore, there is a pressing need to understand how COVID-19 vaccines have been discussed on social media to better understand the public perceptions, concerns, and sentiments that may influence the achievement of herd immunity goals.

Although social media data analysis has been widely performed for both health-related issues and emerging public health crises [10-14], analysis of big data of social media discussion on COVID-19 vaccines is limited [15,16]. To the best of our knowledge, in the most recently published work about COVID-19 vaccine–related social media discussion, the study period ended in November 2020 [16]. However, many significant events related to COVID-19 vaccines have occurred since that date, such as the confirmation of more COVID-19 variant cases in North America by the US Centers for Disease Control and Prevention (CDC), rollout of vaccines, and increased numbers of vaccines showing high efficacy. Previous studies found that changes in social media discussion about vaccine-related topics correspond to the changing reality [15,17]. Thus, research involving recent social media data is needed to fully understand the public discussion on COVID-19 vaccines during the pandemic. In addition, knowledge of the content of COVID-19 vaccine discussion on social media will provide a possible explanation for users' attitudes toward COVID-19 vaccines and COVID-19 vaccine acceptance or hesitancy. However, previous research on COVID-19 vaccines did not touch on these topics [16,18-21]. To fill this gap, in this study, we will examine the public discourse about COVID-19 vaccines on Twitter since the World Health Organization announced it to be a global pandemic on March 11, 2020, up to January 31, 2021, to identify the topics, overarching themes, and sentiments around COVID-19 vaccines and vaccination. This is the first study to include data from almost one year of the pandemic. This long time span will allow us not only to observe a bigger picture of the public discussion and concerns regarding COVID-19 vaccines but also discern the salient changes in major topics and sentiments during the course of the pandemic and further inform public health education and campaigns for increasing COVID-19 vaccine acceptance. Also, our results may provide insight that will be useful in the promotion of other vaccines.

Methods

Data Extraction and Preprocessing

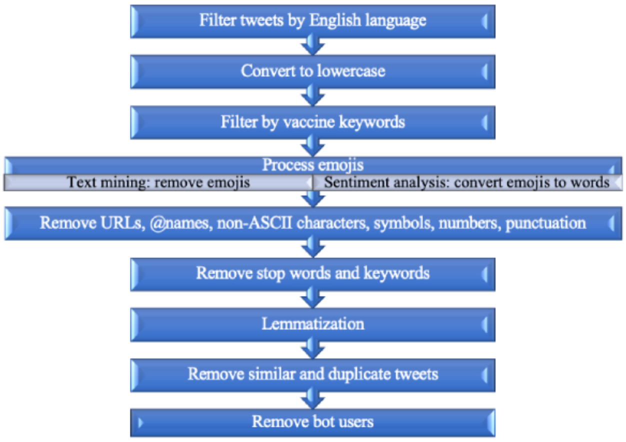
The IDs from a total of 1,499,421 tweets, without retweets, from March 11, 2020, through January 31, 2021, were obtained using the data set maintained by Georgia State University's Panacea Lab [22]. These tweets were collected by the Panacea Lab using the following 13 keywords: *COVID19*, *CoronavirusPandemic*, *COVID-19*, *2019nCoV*, *CoronaOutbreak*, *coronavirus*, *WuhanVirus*, *covid19*, *coronaviruspandemic*, *covid-19*, *2019ncov*, *coronaoutbreak*, and *wuhanvirus*. The Panacea Lab can only provide the tweet IDs [23], which need to be hydrated to recover the full tweet data.

During the tokenizing stage, we used the *gsub* function in R (R Foundation for Statistical Computing) to extract the tweets whose language field in the tweets' metadata was specified as English. All text mining was performed using RStudio Version 1.4.1103 on a Mac computer (Apple Inc) running Big Sur, version 11.2.2. We converted the text of all the tweets to lowercase. We further filtered the tweets by the following keywords: *vaccination*, *vaccinations*, *vaccine*, *vaccines*, *immunization*, *vaccinate*, and *vaccinated*. We prepared 2 batches of tweets, 1 for text mining and 1 for sentiment/emotion analysis. The data processing procedures for the 2 batches are almost the same except at the beginning: for sentiment analysis, we converted all the emojis to words, whereas for text mining, we removed all the emojis. Next, we created a script to remove the URLs, mentioned names, non–American Standard Code for Information Interchange (ASCII) characters, and anything other than English letters or spaces (eg, "1" and "?"). Using the R package *dplyr*, version 1.0.2, we cleaned the tweets by removing duplicates. To filter tweets created by fake or bot accounts, we used the document-term matrix (DTM), which contains rows corresponding to the tweets and columns corresponding to the terms. Each entry in the DTM denotes the number of times a term appears in a tweet. The similarity matrix $S=[S_{ij}]$, which measures how similar the i-th tweet is to the j-th tweet, is obtained by computing the dot product between the i-th row vector and the j-th row vector R_j in the DTM, which geometrically represents the cosine of the angle between the row vectors R_i, R_j:

Therefore, if the i-th and the j-th tweets are identical, then $S_{ij}=1$; if they are completely different (ie, the angle between the corresponding row vectors is 90°), then $S_{ij}=0$. For tweets that were 80% similar, we retained the most representative one (measured by the magnitude of the row vector in the DTM). Furthermore, we used the *tweetornot* package [24], version 0.1.0, to remove users that were identified as bots with a 95% probability or higher.

The final cleaned data set consisted of 1,499,421 unique tweets from 583,499 different users. We further cleaned the tweets by removing words and characters that were of little or no analytical value (eg, “the”, “very”, “&”). We performed this task by creating our own list of stop words by appending the 13 keywords related to *COVID-19* and the 7 keywords related to *vaccine* to the English stop words list from the R package *tidytext*, version 0.2.6; this step was performed because we already knew that every tweet would contain one or more of those keywords, and retaining them in the tweets would not further our understanding of the main content of the tweets. Finally, we stemmed and lemmatized the words to their root forms using the R package *textstem*, version 0.1.4 (eg, *vaccinating*, *vaccinates*, and *vaccinated* were changed to *vaccinate*). Figure 1 shows a summary of our data preprocessing procedure.

Figure 1. Data preprocessing procedure. ASCII: American Standard Code for Information Interchange.



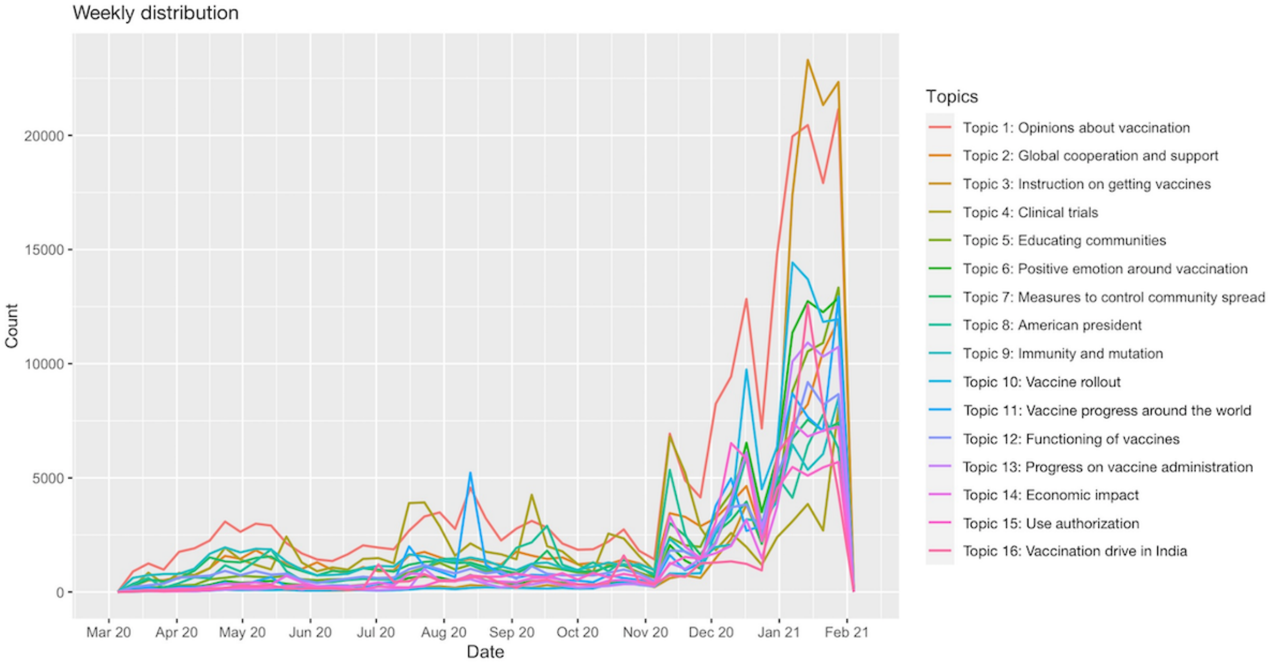
Topic Modeling

Topic modeling can help organize a large collection of documents by grouping documents into different themes. Topic modeling is often referred to as a probabilistic clustering. It is more robust and usually provides more realistic results than hard clustering (eg, k-mean clustering) [25]. A typical clustering algorithm assumes a distance measure between topics and assigns one topic to each document, whereas topic modeling assigns a document to a collection of topics with different weights or probabilities without any assumption of the distance measured between topics. Many topic models are available, of which the most widely used is the latent Dirichlet allocation (LDA) model [25], developed by David Blei, Andrew Ng, and Michael I Jordan in 2002 [26].

To extract common topics from this very high number of tweets, we used the LDA algorithm for topic modeling. We performed the LDA algorithm on the data using the R *textmineR* package, version 3.0.4. The LDA algorithm requires manual input of the number of expected topics. We ran the LDA algorithm on the data by varying the topic number from 2 through 40. For each topic number, we calculated the coherence score using the *textmineR* package. We chose 16 as the topic number for our final topic model, based on two considerations: first, the topic number 16 corresponded to the highest coherence score (see Multimedia Appendix 1); second, in comparison with topics that appeared in the other topic models, the topic model with 16 topics strikes a balance between one too narrow that the model would risk excluding important topics and one so broad that it would dilute the main focus.

The top 8 terms from each of the 16 topics were generated. We also used the *geo_freqpoly* function in the R package *ggplot2*, version 3.3.2, to generate the frequency polygons (see Figure 2) to visualize the weekly frequency of the 16 topics from March 11, 2020, to January 31, 2021. For each tweet, the LDA assigned a probability to each of the 16 topics. We assigned the topic with the highest probability to each tweet, and we grouped the tweets according to the most prevalent topics. To obtain representative tweets for each topic, we randomly sampled 100 tweets from each topic; two of the authors then independently examined the sampled tweets, followed by a group discussion to select the most representative ones. If one of the authors thought that no conspicuous topics had emerged from the first 100 sampled tweets, another 100 tweets would be sampled and further reviewed; the authors continued this process until they judged that there was a clear common topic and they reached a consensus (see our previous paper [27] for more details). We used the *textmineR* package’s topic label function to generate an initial labeling for the topics. After carefully reading through the sampled tweets from each topic, the two authors refined the machine-generated labeling to give each topic the most accurate, concise, and coherent description (see Table 1 in the *Results* section). Through discussions, the authors further grouped the topics into 5 overarching themes. Specifically, two of the authors first independently grouped the topics into the number of themes that made the most sense to them and resolved conflicting views through discussion. The third author was involved by providing additional comments on both the agreements and disagreements between the two authors. The final decision of the grouping was made together by all three authors. For example, whether the topic of “vaccination drive in India” should be grouped into the theme of “vaccine administration” or “vaccines as a global issue” was unresolved after the two-author discussion. By rereading tweets and discussing among the three authors, we finally placed the topic under the theme of vaccines as a global issue.

Figure 2. Weekly frequency of each topic on Twitter from March 11, 2020, to January 31, 2021.



Sentiment and Emotion Analyses

Sentiment analysis can be used to classify the polarity of a given document; it can assign a score to a document to indicate whether the expressed opinion is positive, negative, or neutral. Emotion analysis goes beyond simple polarity and can give scores to different emotions, such as anger, fear, anticipation, trust, surprise, sadness, joy, and disgust (the so-called Plutchik wheel of emotions) [28]. The *syuzhet* package (Jockers, 2017) [29] is the most popular and efficient R package for sentiment/emotion analysis. To perform the emotion analysis, we used the National Research Council of Canada Emotion Lexicon developed by Turney and Mohammad in 2010 [30]. It is the most comprehensive dictionary for this task [31]. In Figure 3, we show the weekly average polarity (sentiment) scores from March 11, 2020, to January 31, 2021; we fitted the data points with the best linear fit and obtained a slope of 0.003764 with a *P* value <.001 and an intercept of 0.1653927 with a *P* value <.001. In Figure 4, we plot the weekly percentage of emotions.

Figure 3. Weekly average polarity (sentiment) scores from March 11, 2020, to January 31, 2021. The slope of the best fit is 0.003764.

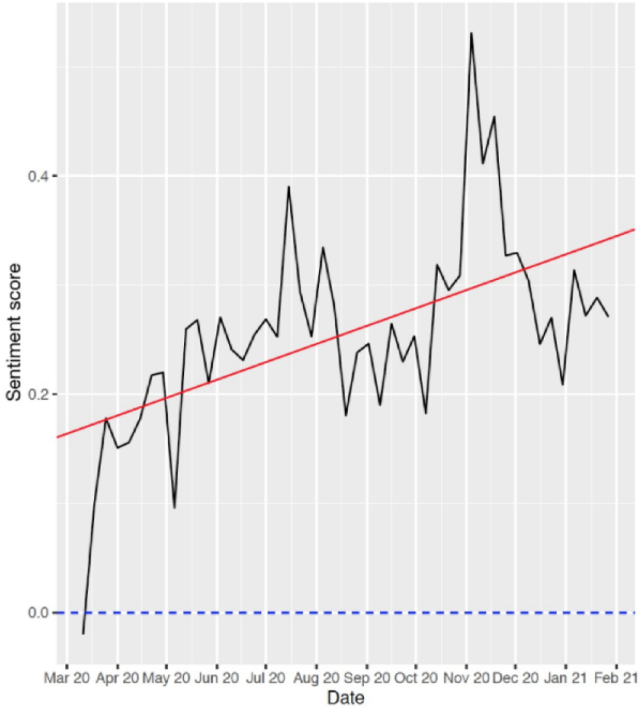
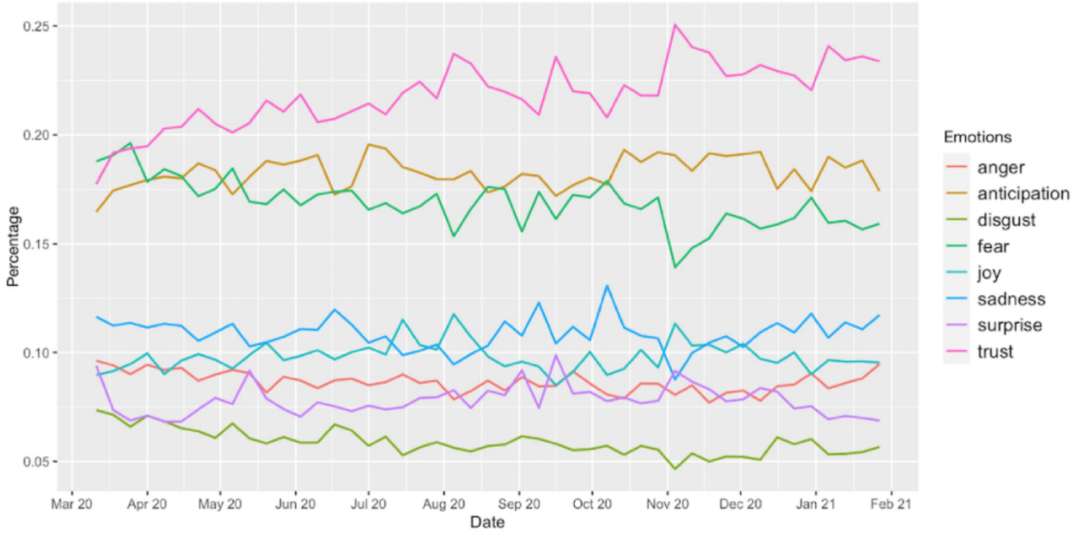


Figure 4. Weekly percentages of emotions from March 11, 2020, to January 31, 2021.

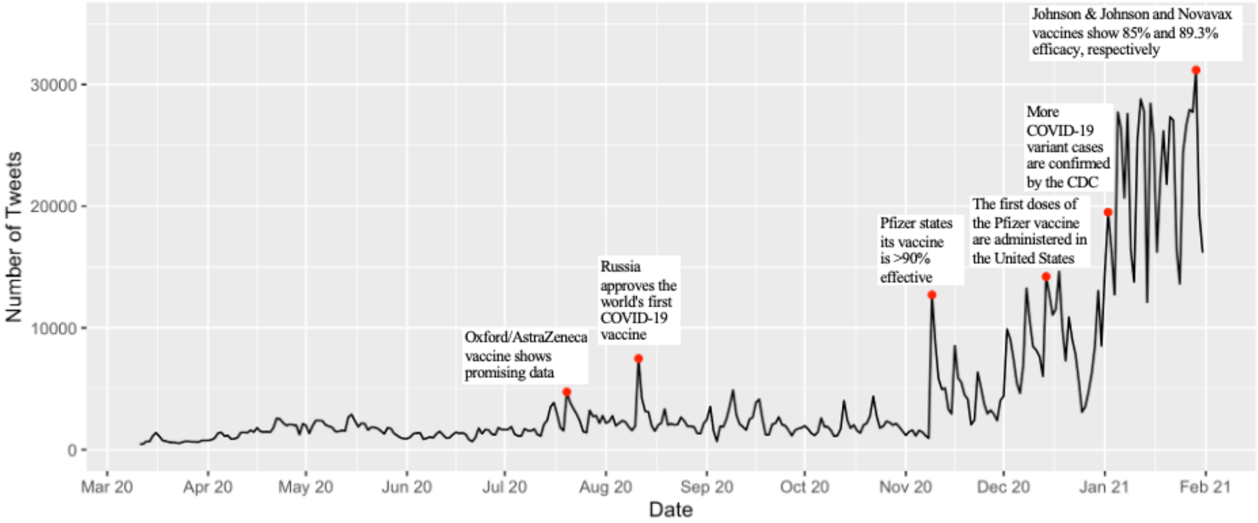


Results

Numbers of Tweets

We downloaded 144,332,894 tweets posted from March 11, 2020, through January 31, 2021 (for a total of 327 days) using the Panacea Lab database. After cleaning, a total of 1,499,421 tweets from 583,499 different users were included in the analysis. As shown in Figure 5, the number of daily tweets continually increased; the daily average for the month of January 2021 was 22,202 tweets. Before November 9, 2020, the number of daily tweets was around or below 5000, with only one exception, on August 11, 2020 (n=7486), when Russia approved the world's first COVID-19 vaccine [32]. The first wave of an exponential increase in the number of daily tweets started on November 9, 2020 (n=12,720), when Pfizer stated that its vaccine is 90% effective [33]. The second wave of a surge in the number of daily tweets started around January 3, 2021, when more COVID-19 variant cases were confirmed in North America by the CDC, and the highest number of tweets in a single day (n=31,197) occurred on January 29, 2021, when Johnson & Johnson's and Novavax's vaccines showed 85% and 89.3% efficacy, respectively [34].

Figure 5. Daily numbers of COVID-19–related tweets from March 11, 2020, to January 31, 2021.



Topic Modeling of COVID-19 Vaccine–Related Tweets

Analysis of the tweets yielded 16 topics, which were grouped into 5 overarching themes. In descending order of quantity of tweets (N=1,499,421), the themes are opinions and emotions around vaccines and vaccination (405,397 tweets, 27.04%), knowledge around vaccines and vaccination (355,305 tweets, 23.7%), vaccines as a global issue (311,251 tweets, 20.76%), vaccine administration (266,744 tweets, 17.79%), and progress on vaccine development and authorization (160,724 tweets, 10.72%). Table 1 summarizes the 16 topics, top terms in the topics, and the number and percentage of each topic; it also provides a tweet example for each topic. More details about the themes and topics, including the salient temporal variance of topics (see Figure 2), are elaborated in the following subsections.

Table 1. Numbers and percentages of tweets related to each COVID-19 vaccine topic (N=1,499,421).

Theme and topics	Top terms contributing to the topic model	Total tweets, n (%)	Example paraphrased tweet[a] (date posted)
Theme 1: Opinions and emotions around vaccines and vaccination			
Opinions about vaccination	*people*, *get*, *take*, *go*, *want*, *make*, *like*, *think*	227,840 (15.20)	"It is pathetic to see the low trust in science and the government. People need to see leaders and politicians on TV to receive the vaccine to convince them that the vaccine is safe." (December 18, 2020)
Positive emotion around vaccination	*get*, *good*, *first*, *day*, *today*, *work*, *one*, *feel*	89,117 (5.94)	"After getting the first dose of the covid-19 vaccine today, I can finally breath some fresh air and feel there is hope in life." (January 23, 2021)
American president	*Trump*, *Biden*, *American*, *president*, *gate*, *bill*, *plan*, *administration*	88,440 (5.90)	"There are more lies than usual in today's press. Biden is being portraited as anti-vaccine. This is not correct. In fact, he said on Wednesday he trusted the vaccines and the scientists, while accusing President Trump of playing politics with a potential covid-19 vaccine." (September 16, 2020)
Theme 2: Knowledge about vaccines			
Educating communities	*question*, *read*, *answer*, *expert*, *article*, *community*, *immunity*, *black*	96,532 (6.44)	"You can learn more about the covid-19 vaccines in our upcoming town hall meeting. We will address any questions and concerns on January 13, 2021, 2-3:30 PM (EST)." (January 7, 2021)
Measures to control community spread	*mask*, *get*, *need*, *go*, *still*, *available*, *end*, *life*	89,008 (5.94)	"Even though covid-19 vaccines are an important step in slowing down the spread of the virus, people still need to continue taking all precautions: wear a mask, maintain physical distance from others, and keep your hands clean." (January 19, 2021)
Immunity and mutation	*flu*, *year*, *new*, *may*, *variant*, *work*, *effective*, *strain*	88,327 (5.89)	"They say that coronavirus resembles the flu virus. As we know, the flu virus mutates. Therefore, they need to create different flu shot every year into order to fight off the virus, but there hasn't been a flu vaccine that is 100% effective. So good luck with making the covid-19 vaccine." (May 11, 2020)
Functioning of vaccines	*people*, *effect*, *test*, *risk*, *side*, *side effect*, *immune*, *death*	81,438 (5.43)	"Both vaccines use mRNA technology which contains instructions that tell our cells how to make a protein that triggers an immune response inside our bodies." (December 21, 2020)
Theme 3: Vaccines as a global issue			
Global cooperation and support	*world*, *country*, *global*, *pandemic*, *access*, *need*, *develop*, *effort*	108,366 (7.23)	"On Thursday, WHO [the World Health Organization] warns against 'vaccine nationalism.' No one country is safe if poor countries can't get access to the vaccines." (August, 9, 2020)
Vaccine progress around the world	*UK*, *China*, *EU*, *Russia*, *first*, *country*, *approve*, *news*	83,156 (5.55)	"On Tuesday, German health minister Jens Spahn joined EU [the European Union] to place restriction on exporting covid-19 vaccines. This came amid discontent over the delay in rolling out the AstraZeneca vaccine to member countries." (January 26, 2021)
Economic impact	*new*, *case*, *death*, *news*, *rollout*, *rate*, *high*, *report*	61,360 (4.09)	"A rise in new covid-19 cases outweigh positive vaccine news. Shutdown fears sink global markets. US coronavirus deaths are above 250,000." (November 9, 2020)
Vaccination drive in India	*India*, *drive*, *minister*, *Indian*, *speed*, *warp*, *health*, *warp speed*	58,369 (3.89)	"India clears covid-19 vaccine makers Serum Institute of India (SII) and Bharat Biotech to start the world's largest immunization drive." (January 3, 2021)
Theme 4: Vaccine administration			
Instruction on getting vaccines	*county*, *health*, *state*, *appointment*, *update*, *plan*, *site*, *distribution*	106,544 (7.11)	"Health departments will each announce their vaccine availability and locations. Eligible individuals at the federally qualified health center will be contacted regarding access to a covid-19 vaccine." (December 30, 2020)

Theme and topics	Top terms contributing to the topic model	Total tweets, n (%)	Example paraphrased tweet[a] (date posted)
Vaccine rollout	*worker*, *care*, *health*, *first*, *receive*, *healthcare*, *hospital*, *staff*	85,623 (5.71)	"We have many sites delivering the covid-19 vaccines to the top priority group. Please be patient and if you are in the top priority group, you will be contacted." (January 17, 2021)
Progress on vaccine administration	*dose*, *million*, *first*, *receive*, *week*, *Pfizer*, *people*, *first dose*	74,577 (4.97)	"Since Tuesday's update, Louisiana has administered 25,133 additional covid-19 vaccines. The second doses started being administered this week. 7,068 people have been fully vaccinated. In total, since mid-December, 81,585 doses have been administered." (January 7, 2021)
Theme 5: Progress on vaccine development and authorization			
Clinical trials	*trial*, *clinical*, *Pfizer*, *Moderna*, *phase*, *effective*, *clinical trial*, *oxford*	99,754 (6.65)	"There are 70 covid-vaccines under development. Moderna is one of the first to test their covid-19 vaccine on humans. If the trial is successful, Moderna could reach the final stage of testing by Fall 2020." (April 26, 2020)
Use authorization	*use*, *FDA*, *approve*, *emergency*, *approval*, *Pfizer*, *emergencyuse*, *health*	60,970 (4.07)	"The FDA's [US Food and Drug Administration's] vaccine advisory committee has unanimously approved an emergency use of the covid-19 vaccine made by Pfizer and by Moderna." (January 19, 2021)

[a] The tweets are paraphrased to protect users' privacy.

Theme 1: Opinions and Emotions Around Vaccines and Vaccination

This theme contained 3 topics, among which the topic of opinions about vaccination (227,840/1,499,421 tweets, 15.2%) was the most tweeted of all the 16 topics and remained the most discussed topic in the majority of the period of our examination (see Figure 2). This theme reflected the mixed opinions of Twitter users about vaccination, including their doubt, hesitancy, trust, and advocacy. Some Twitter users also asked for freedom about vaccination. The second topic, positive emotion around vaccination, featured happiness, hope, relief and other positive emotions that the public showed in their tweets. Many emotions were based on their direct or indirect experiences with vaccination. The third topic was named "American president." The public's opinions and emotions were expressed through their comments on US President Trump and President Biden's vaccine-related words and actions.

Theme 2: Knowledge Around Vaccines and Vaccination

This theme consisted of 4 topics focusing on understanding and facilitating understanding of COVID-19 vaccines and vaccination. The most frequently discussed topic under this theme was educating communities, with a considerable number of tweets spreading information about the live webinars where health professionals would provide important information about COVID-19 vaccines and answer questions. The topic of measures to control community spread featured the recommendation of taking measures such as wearing a mask, washing hands, and social distancing both before and after getting vaccinated. The topics of immunity and mutation centered around the immunity of vaccines and mutation of the coronavirus, which was frequently discussed in comparison with *flu virus*. Functioning of vaccines, as a topic, introduced how vaccines work and the subsequent symptoms and side effects after vaccination.

Theme 3: Vaccines as a Global Issue

The theme of vaccines as a global issue consisted of 4 topics related to the globality of SARS-CoV-2 and the COVID-19 vaccines; therefore, the topics under this theme were not centered on the United States. These topics involved active news about vaccine progress and updates around the world, such as international vaccine supplies, delivery, and purchases in many countries (ie, the topic of vaccine progress around the world), the impact of vaccines on the global economy (ie, the topic of economic impact), and the world's largest inoculation drive (ie, the topic of the vaccination drive in India). The topic of vaccine progress around the world became the most discussed topic around August 11, 2020, because Russia approved the world's first COVID-19 vaccine (see Figure 2). The most salient topic under this theme was global cooperation and support, which called for global cooperation to accelerate the vaccine development and equitable access and advocated no vaccination nationalism. This was also the second most tweeted topic (108,366/1,499,421 tweets, 7.23%) among the 16 topics, only after the topic of opinions about vaccination.

Theme 4: Vaccine Administration

This theme consisted of 3 topics covering several aspects of vaccine administration. The topic of vaccine rollout mainly centered around vaccination to the top priority groups, including health care providers being the first in line. After the first doses of Pfizer vaccine were administered in the United States on December 14, this topic remained one of the 3 most discussed topics for approximately 5 weeks (see Figure 2). The second topic focused on the progress of the vaccine administration, including the shipment and supply of vaccines that directly relate to the progression of vaccine administration. The third topic under this theme was instruction on obtaining vaccines, which included spreading of information from health authorities at various levels to guide the public to obtain their vaccine shots. More detailed information was mentioned in the tweets in this topic, such as “If anyone is 75 and older, does not have internet access, and needs help to schedule an appointment, please call the central appointment desk for help at 1-866-960-0633” (a tweet posted on January 20, 2021). With the advancement of vaccine administration, this topic gradually became more salient and leaped to the rank of most discussed topic, surpassing the topic of opinions about vaccination, after the first week of January; it then remained at the top rank until the end of January (see Figure 2).

Theme 5: Progress on Vaccine Development and Authorization

There are 2 topics under this theme, which focused on the development of COVID-19 vaccines and authorization by the US Food and Drug Administration (FDA). The topic of clinical trials centered around the plan and process of clinical trials, mainly from Pfizer and Moderna, and the updated results of the clinical trials. This topic remained the most discussed topic for a week around July 20, when the first human trial of the Oxford/AstraZeneca coronavirus vaccine showed promise [35], and became 1 of the 2 most discussed topics, together with the topic of opinions about vaccination, for 3 weeks after Pfizer stated its vaccine is 90% effective on November 9, 2020 (see Figure 2). The topic of use authorization centered on FDA approval of an emergency use authorization for COVID-19 vaccines.

Sentiment and Emotion Analyses of COVID-19 Vaccine Tweets

Weekly mean sentiment scores showed that despite fluctuations, in general, the sentiment was increasingly positive from March 11, 2020, to January 31, 2021, with the linear best fit slope of 0.003764 (with a *P* value <.001, which indicates that it is statistically significant) (see Figure 3). Moreover, the positive emotion reached its peak around November 9, 2020, when Pfizer announced that its vaccine is 90% effective; on the same day, the number of daily tweets became historically high before January, as mentioned above. Emotion analysis further showed that trust was the most predominant emotion, accounting for 22.78% of the 8 emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust), followed by anticipation (18.34%), fear (16.29%), sadness (10.97%), joy (9.76%), anger (8.63%), surprise (7.60%), and disgust (5.63%). Noticeably, the most dominant emotion shown in COVID-19 vaccine tweets before April was fear; however, it changed to trust from the week of April 1, 2020 (Figure 4). Trust remained the most dominant emotion after that, and the number of tweets expressing trust continued to grow. In addition, it was observed that when the emotion of trust increased, the emotion of fear decreased. The trust emotion reached its peak on November 9, 2020, when Pfizer announced that its vaccine is 90% effective; on the same day, the fear emotion was expressed the least throughout the time period of our examination. It was also noted that apart from the obvious changes in the emotions of trust and fear over time, the other emotions were relatively stable in the period of March 11, 2020, to January 31, 2021 (see Figure 4).

Discussion

Principal Findings

In this study, we examined sentiments and topics over a long time span, covering the discussions about COVID-19 vaccines from when COVID-19 became a global pandemic (March 11, 2020) to January 31, 2021, when multiple vaccines become available and mass vaccination had begun in the United States and many other countries. This study adds to the latest research about the social impact of the COVID-19 vaccine. For example, there are surveys addressing sociodemographic social media user characteristics and the social determinants of vaccine acceptance [18,36]. Our research supplements studies like these by providing the discourse patterns on social media, that is, how people actually talked about their vaccination intentions and other related issues. Researchers have taken a similar approach to studying very specific vaccine-related topics, such as side effects and the type of vaccines in the context of China. This study could provide a comparison in different cultural contexts, as the corpus was pulled from a global data set and English tweets were analyzed. The specific topics and public sentiments identified could be used for further studies about specific vaccine-related topics among English language social media users. Many of the current studies about COVID-19 vaccination focus on vaccine hesitancy and antivaccine messages [19], and some used survey methods to gather data [37]. As the vaccines for COVID-19 are still very new, and they were developed in a very short period of time due to the urgent need, we would expect the expression of vaccine hesitancy in the public media channels. This study would help contextualize the major concerns about vaccine efficacy and safety.

This study has found some changing patterns of the discussion on Twitter along with the progress. Similar patterns of sentiments and topics found in another research using natural language processing and deep learning techniques on Facebook and Twitter posts [16]. The results of this research show that the patterns are valid across platforms. The number of tweets regarding COVID-19 vaccination is largely driven by major events—mainly the milestones in vaccine development and the new variants of the virus. Major spikes in the number of tweets correspond closely to these events. The analysis of each topic also shows such patterns. For example, from the weekly distribution of topics, a sudden overall increase is noticeable starting from early November, also right at the time of the Pfizer announcement. Scholars have studied how information flows from social media to mainstream news. In today’s media ecology, the boundary between social and mainstream media is no longer clear; however, here we can see that social media discourse is largely mirroring what is happening in the news. Future studies could detect to what extent the social media discussion is shaped by mainstream media.

The sentiment analysis shows that the general sentiment toward COVID-19 vaccination is becoming more positive over time. The overall sentiment score reached the maximum in early November 2020, which also corresponded to the report of the high efficacy of the Pfizer vaccine. As for the emotions, trust continually dominated the discussion; it reached the highest point around early November, following Pfizer's announcement about the efficacy of its vaccine. The percentage of tweets increased overall, showing that more people were expressing their trust in the discussion of COVID-19 vaccine. The change of the percentage of tweets expressing fear mirrors that of tweets expressing trust. The overall percentage is decreasing, showing that as the vaccine development progressed, people's fear about the pandemic decreased. The highest point was in mid-March 2020, after the declaration of the global pandemic, and the lowest point was in early November, at the time of Pfizer's announcement. As vaccine research and testing moved closer to a promising result, the expression of fear declined. Other emotions, in terms of their percentages of the overall tweets, remained more or less stable over time. Trust is the dominant emotion, which can be understood as a reflection of the vaccination as the only option. Unlike other types of vaccines, which people can choose to take or not, vaccination has been increasingly viewed as the only promising way to end the pandemic given the prevalence of COVID-19, the speed of its community spread, the disruption of normal life, and the lack of other options proved to be efficient.

The topics related to opinions and emotions were the most common, and among the 3 topics in this theme, the topic on opinions about vaccination represented the largest proportion. The development of COVID-19 vaccines has progressed along with the spread of the virus and the appearance of variants, together with our increasing knowledge about the disease, all of which have become active topics in public discourse. With many uncertainties remaining about vaccines and other options, we would expect mixed opinions to surface on the platform. This mixed opinion about vaccination should also be situated in the larger context of the antivaccine movement in the United States and other countries as well. Our results show that doubts, vaccine hesitancy, conspiracy theories, and the argument that vaccination is an individual freedom are all common themes in the antivaccine discourse. However, as mentioned earlier, as COVID-19 has affected people's day-to-day life, a vaccine is crucial for returning to normal life; people still have much hope in vaccines, leading to the sharing of positive emotions. Finally, with the political climate and the ongoing presidential election, both candidates included vaccination on their agenda; thus, the discussion about the vaccine was very politicized.

As COVID-19 is an ongoing crisis with a global scale, the discussion about the vaccine is also global. The pandemic has revealed how much the world is now connected, and thus vaccination has become a global issue—if a country cannot reach a certain level of vaccination of its population, it has a high risk of contagion and virus mutation; thus, it is difficult for the country to return to its role in the global economy, and global cooperation is needed to defeat the disease. Therefore, the economic impact of the pandemic and the development of vaccines are salient issues.

The currently available vaccines are also results of cross-national collaboration, which is why "vaccine nationalism" is frequently mentioned in tweets as an issue that would impede the progress of fighting the pandemic. In a global crisis, information-seeking needs increase, particularly as vaccines become available and people need instructive information regarding the vaccine rollout, obtaining vaccines, and the vaccine administration. As COVID-19 is a new disease that is still being studied, and the vaccines also became available after a short period of research and development, there are numerous uncertainties facing public acceptance of these vaccines. For laypeople, the science behind the vaccines is still not well understood, particularly because COVID-19 is such a new disease and many things are still unknown. These uncertainties have provoked people to seek and share information about the vaccines, which is reflected in the topic of knowledge around vaccines, such as the science behind the vaccines (immunity and mutation), the techniques used, and results of clinical trials. These uncertainties have also provoked people to seek information about the vaccine rollout and administration because these topics are are closely connected to when and how people will be vaccinated.

The results of this study show that the discussions about COVID-19 vaccines are multifaceted, and the public are actively seeking and sharing information about them. It is important for public health agencies to understand the major public interests and concerns regarding vaccination, that is, the major factors that would affect vaccine acceptance and hesitancy. In this way, they could establish appropriate strategies to facilitate public communication. Our study shows that the major spikes in vaccine discussions corresponded to reports of the breakthroughs of vaccine development as they appeared in the news. Public health agencies need to pay attention to this pattern and monitor the discussions on the web on days when major news is reported about vaccine development or other significant events, such as a report of severe adverse events from a particular vaccine. These agencies could develop immediate responses based on identifying the instant reactions—and the dominant emotions and topics—on social media. In this study, we found that trust is the predominant emotion regarding vaccines, which is promising and reassuring for the public health agencies promoting vaccination. For individual themes and topics, public health agencies should pay close attention to the discussions about vaccine knowledge and administration to close the information gap between the needs of the public and the information the agencies have provided.

Limitations

By examining how topics and sentiments evolve along the timeline of the pandemic and vaccine development, one can see the correspondence of the volume of tweets and major research breakthroughs in the news. Although it is clear that the number of tweets shows a major spike at the news of Pfizer reporting its high vaccine efficacy in early November, a statistical analysis may provide more details about whether there are significant differences. Although it is not the focus of this study, close examination of Twitter users may provide more meaningful information, such as how the contents shared by different kinds of users vary and whether certain types of users are more likely to post or comment on certain topics. In addition, this study may have geographic bias in the examined tweets, as is observed in many other studies of unstructured textual data [38]. As a result, a given data set could overrepresent some geographic areas. Last but not least, because Twitter users are not representative of the US population [39], our data set could be overly representing a subset of the population with specific characteristics. Therefore, the findings of this study should be generalized with great caution.

Conclusion

This study identifies the major topics and sentiments about COVID-19 vaccine–related issues discussed on social media. It also examines the changes in these topics and sentiments over time to better understand the larger trend. Among the 16 distinct topics, opinions about vaccination was the most common and remained so over time. As vaccine development progressed around the world, the dominant topics also shifted. Instructions on getting the vaccine became the most discussed topic around early January 2021. The discussion of COVID-19 vaccination on social media was largely driven by major news events about COVID-19 vaccines and mirrored the active new topics in mainstream media. Also, the discussion has a global perspective. The overall sentiment was increasingly positive over time, and trust was the predominant emotion, which shows that social media discussions may imply higher acceptance of COVID-19 vaccines compared with previous vaccines. Due to the timeline of our data set, in this study, we did not further examine the sentiments about specific brands of vaccines. We would expect the discussions to be different when using different brand names to search for tweets and conduct sentiment analysis. Particularly, after administration of the Johnson & Johnson vaccine was paused by the CDC, there could be a surge of related discussions, and a topic about the side effects of the vaccines could emerge. Therefore, further study in this line is highly recommended.

Acknowledgments

This research was supported by National Cancer Institute Grant T32 CA 113710 (JL).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Coherence scores of the different numbers of topics.
[DOCX File , 1098 KB - jmir_v23i6e24435_app1.docx]

References

1. Le TT, Cramer JP, Chen R, Mayhew S. Evolution of the COVID-19 vaccine development landscape. Nat Rev Drug Discov 2020 Oct 04;19(10):667-668. [doi: 10.1038/d41573-020-00151-8] [Medline: 32887942]
2. Gottlieb S. America needs to win the coronavirus vaccine race. The Wall Street Journal. 2020 Apr 26. URL: https://www.wsj.com/articles/america-needs-to-win-the-coronavirus-vaccine-race-11587924258 [accessed 2021-04-09]
3. Andre FE, Booy R, Bock HL, Clemens J, Datta SK, John TJ, et al. Vaccination greatly reduces disease, disability, death and inequity worldwide. Bull World Health Organ 2008 Feb;86(2):140-146 [FREE Full text] [doi: 10.2471/blt.07.040089] [Medline: 18297169]
4. Abbasi J. COVID-19 conspiracies and beyond: how physicians can deal with patients' misinformation. JAMA 2021 Jan 19;325(3):208-210. [doi: 10.1001/jama.2020.22018] [Medline: 33377900]
5. Ball P. Anti-vaccine movement could undermine efforts to end coronavirus pandemic, researchers warn. Nature 2020 May 13;581(7808):251-251. [doi: 10.1038/d41586-020-01423-4] [Medline: 32405043]
6. Kwok KO, Lai F, Wei WI, Wong SYS, Tang JW. Herd immunity - estimating the level required to halt the COVID-19 epidemics in affected countries. J Infect 2020 Jun;80(6):e32-e33 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.027] [Medline: 32209383]
7. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 2020 Jul;26(7):1470-1477 [FREE Full text] [doi: 10.3201/eid2607.200282] [Medline: 32255761]
8. Tyson A, Johnson C, Funk C. U.S. public now divided over whether to get COVID-19 vaccine. Pew Research Center. 2020 Sep 17. URL: https://www.pewresearch.org/science/2020/09/17/u-s-public-now-divided-over-whether-to-get-covid-19-vaccine/ [accessed 2021-04-15]
9. Ashkenazi S, Livni G, Klein A, Kremer N, Havlin A, Berkowitz O. The relationship between parental source of information and knowledge about measles / measles vaccine and vaccine hesitancy. Vaccine 2020 Oct 27;38(46):7292-7298. [doi: 10.1016/j.vaccine.2020.09.044] [Medline: 32981777]
10. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. J Med Internet Res 2015 Oct 27;17(10):e243. [doi: 10.2196/jmir.4969] [Medline: 26508089]
11. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. Am J Infect Control 2010 Apr;38(3):182-188 [FREE Full text] [doi: 10.1016/j.ajic.2009.11.004] [Medline: 20347636]

12. Lazard AJ, Scheinfeld E, Bernhardt JM, Wilcox GB, Suran M. Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. Am J Infect Control 2015 Oct 01;43(10):1109-1111. [doi: 10.1016/j.ajic.2015.05.025] [Medline: 26138998]
13. Masri S, Jia J, Li C, Zhou G, Lee M, Yan G, et al. Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic. BMC Public Health 2019 Jun 14;19(1):761 [FREE Full text] [doi: 10.1186/s12889-019-7103-8] [Medline: 31200692]
14. Signorini A, Segre A, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
15. Bonnevie E, Gallegos-Jeffrey A, Goldbarg J, Byrd B, Smyser J. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. J Commun Healthc 2020 Dec 15;14(1):12-19. [doi: 10.1080/17538068.2020.1858222]
16. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. J Med Internet Res 2021 Apr 05;23(4):e26627 [FREE Full text] [doi: 10.2196/26627] [Medline: 33724919]
17. 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]
18. Benis A, Seidmann A, Ashkenazi S. Reasons for taking the COVID-19 vaccine by US social media users. Vaccines (Basel) 2021 Mar 29;9(4):315 [FREE Full text] [doi: 10.3390/vaccines9040315] [Medline: 33805283]
19. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593. [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
20. Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. EClinicalMedicine 2020 Sep;26:100495 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100495] [Medline: 32838242]
21. 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 05;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
22. Banda J, Tekumalla R, Wang G, Yu J, Liu T, Ding Y. A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration (Version 32) Data set. Zenodo. 2020 Oct 18. URL: http://doi.org/10.5281/zenodo.4104587 [accessed 2021-01-31]
23. The Panacea Lab: COVID-19 Twitter. GitHub. URL: https://github.com/thepanacealab/covid19_twitter [accessed 2021-02-02]
24. tweetbotornot. GitHub. URL: https://github.com/mkearney/tweetbotornot [accessed 2021-06-22]
25. Blum A, Hopcroft J, Kannan R. Foundations of Data Science. Cambridge, UK: Cambridge University Press; 2020.
26. Blei D, Ng A, Jordan M. Latent Dirichlet allocation. J Mach Learn Res 2003 Mar:993-1022 [FREE Full text]
27. Lyu JC, Luli GK. Understanding the public discussion about the Centers for Disease Control and Prevention during the COVID-19 pandemic using Twitter data: text mining analysis study. J Med Internet Res 2021 Feb 09;23(2):e25108 [FREE Full text] [doi: 10.2196/25108] [Medline: 33497351]
28. Plutchik R. A general psychoevolutionary theory of emotion. In: Robert P, Henry K, editors. Theories of Emotion. Cambridge, MA: Academic Press; 1980:3-33.
29. Misuraca M, Alessia F, Germana S, Maria S. Sentiment Analysis for Education with R: packages, methods and practical applications. ArXiv. Preprint posted online on May 08, 2020 [FREE Full text]
30. Turney P, Mohammad S. Emotions evoked by common words and phrases: using Mechanical Turk to create an emotion lexicon. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. 2010 Jun Presented at: NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text; June 2010; Los Angeles, CA p. 26-34 URL: https://www.aclweb.org/anthology/W10-0204
31. Naldi M. A review of sentiment computation methods with R packages. ArXiv. Preprint posted online on January 24, 2019 [FREE Full text]
32. Akst J. Russia approves world's first coronavirus vaccine. The Scientist. 2020 Aug 11. URL: https://www.the-scientist.com/news-opinion/russia-approves-worlds-first-coronavirus-vaccine-67810 [accessed 2021-06-22]
33. Pfizer and Biontech announce vaccine candidate against COVID-19 achieved success in first interim analysis from phase 3 study. Pfizer. 2020 Nov 09. URL: https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against [accessed 2022-06-22]
34. Callaway E, Mallapaty S. Novavax offers first evidence that COVID vaccines protect people against variants. Nature 2021 Jan 29;590(7844):17. [doi: 10.1038/d41586-021-00268-9] [Medline: 33510489]
35. First human trial of Oxford coronavirus vaccine shows promise. Reuters. 2020 Jul 20. URL: https://www.reuters.com/article/us-health-coronavirus-oxford-vaccine-idUSKCN24L1MP [accessed 2020-06-22]
36. Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. EClinicalMedicine 2020 Sep;26:100495 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100495] [Medline: 32838242]

37. Alley S, Stanton R, Browne M, To Q, Khalesi S, Williams S, et al. As the pandemic progresses, how does willingness to vaccinate against COVID-19 evolve? Int J Environ Res Public Health 2021 Jan 19;18(2) [FREE Full text] [doi: 10.3390/ijerph18020797] [Medline: 33477825]
38. 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]
39. 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 2021-06-22]

Abbreviations

ASCII: American Standard Code for Information Interchange
CDC: US Centers for Disease Control and Prevention
DTM: document-term matrix
FDA: US Food and Drug Administration
JSON: JavaScript Object Notation
LDA: latent Dirichlet allocation

Edited by C Basch; submitted 18.04.21; peer-reviewed by J Turner, A Dormanesh, R Gore; comments to author 22.05.21; revised version received 10.06.21; accepted 10.06.21; published 29.06.21.

Please cite as:
Lyu JC, Han EL, Luli GK
COVID-19 Vaccine–Related Discussion on Twitter: Topic Modeling and Sentiment Analysis
J Med Internet Res 2021;23(6):e24435
URL: https://www.jmir.org/2021/6/e24435
doi:10.2196/24435
PMID:34115608

Viewpoint

Developing Digital Tools for Remote Clinical Research: How to Evaluate the Validity and Practicality of Active Assessments in Field Settings

Jennifer Ferrar[1,2], PhD; Gareth J Griffith[2,3], PhD; Caroline Skirrow[1,4], PhD; Nathan Cashdollar[4,5], PhD; Nick Taptiklis[4], BA; James Dobson[4], MEng; Fiona Cree[4], BSc; Francesca K Cormack[4], PhD; Jennifer H Barnett[4,6], PhD; Marcus R Munafò[1,2], PhD

[1]School of Psychological Science, Faculty of Life Sciences, University of Bristol, Bristol, United Kingdom
[2]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
[3]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
[4]Cambridge Cognition Ltd, Cambridge, United Kingdom
[5]Cambridge Cognition Ltd, Cambridge, MA, United States
[6]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Jennifer Ferrar, PhD
School of Psychological Science
Faculty of Life Sciences
University of Bristol
12a Priory Road
Bristol, BS8 1TU
United Kingdom
Phone: 44 117 928 9707
Email: jennifer.ferrar@bristol.ac.uk

Abstract

The ability of remote research tools to collect granular, high-frequency data on symptoms and digital biomarkers is an important strength because it circumvents many limitations of traditional clinical trials and improves the ability to capture clinically relevant data. This approach allows researchers to capture more robust baselines and derive novel phenotypes for improved precision in diagnosis and accuracy in outcomes. The process for developing these tools however is complex because data need to be collected at a frequency that is meaningful but not burdensome for the participant or patient. Furthermore, traditional techniques, which rely on fixed conditions to validate assessments, may be inappropriate for validating tools that are designed to capture data under flexible conditions. This paper discusses the process for determining whether a digital assessment is suitable for remote research and offers suggestions on how to validate these novel tools.

(J Med Internet Res 2021;23(6):e26004) doi:10.2196/26004

KEYWORDS

digital assessment; remote research; measurement validity; clinical outcomes; ecological momentary assessment; mobile phone

Introduction

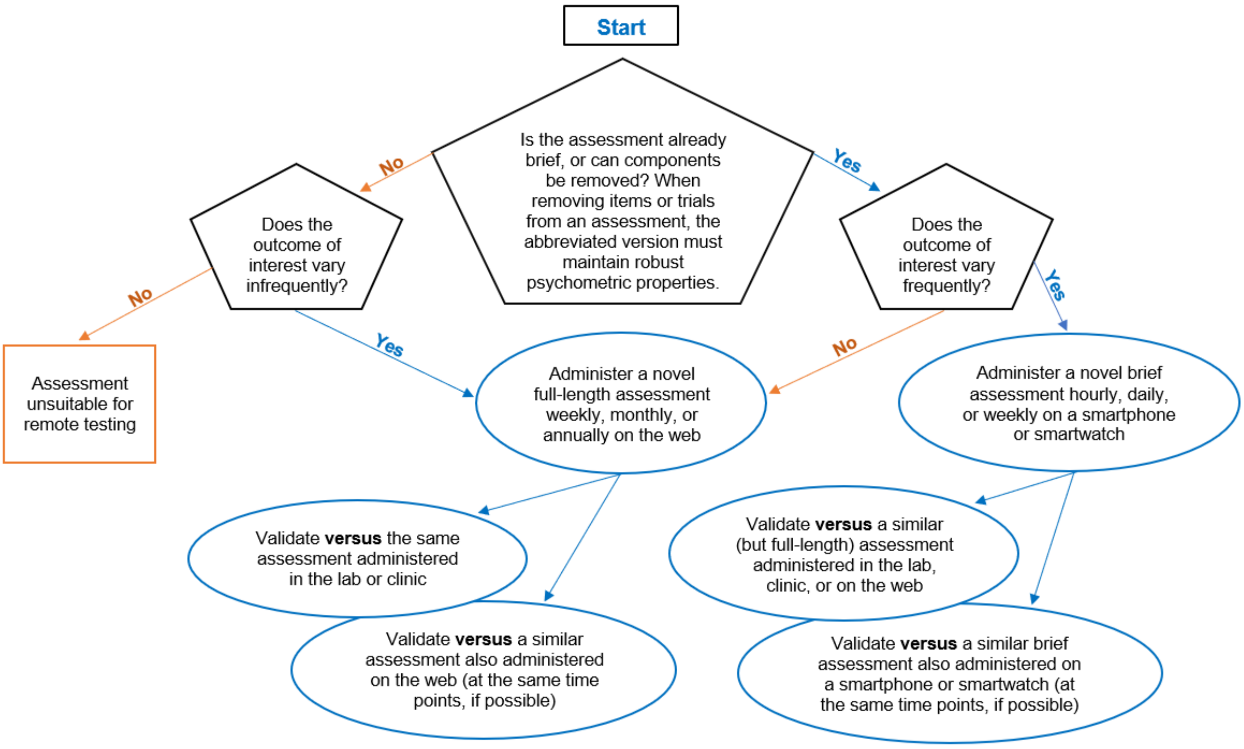
The emergence of SARS-CoV-2 at the turn of 2020 demonstrates how abruptly life—and research—can change. The global response to the resulting pandemic also demonstrates how quickly the world can use technology to adapt to these changes. The physical closure of organizations has less impact now than it would have had even 10 years ago; thanks to technological advances, many formerly in-person activities can now be conducted virtually. For some organizations, this way of operating was already familiar, whereas for others it is novel and challenging. Overall, most organizations are being compelled to adapt and create innovative ways to enhance remote working.

Scientific research has also had to adapt to these unforeseen circumstances. Fortunately, a great deal of psychological research was routinely conducted remotely before the SARS-CoV-2 outbreak [1], primarily in an attempt to produce more externally valid research [2,3]. Remote data collection offers an opportunity for researchers to broaden the diversity of their samples, both in terms of whom they recruit and when and where data collection occurs. Remote data collection is facilitated through web-based recruitment platforms (eg, *Amazon Mechanical Turk*, *Prolific*, and *Call for Participants*), web-based survey and experiment builders (eg, *Qualtrics* and *Gorilla*), and personal devices (eg, smartphones and smartwatches) that can be used to collect data on a range of behaviors (eg, location, movement, social interactions, travel behavior, energy intake, energy expenditure, vital signs, sleep patterns, menstrual cycles, mood, cognition, and pain) [1,3,4]. Smart devices are used for both active and passive data collection, and users can manually input self-report data, whereas built-in sensors allow for continuous collection of objective data [3].

Although much cognitive and behavioral research was already moving toward remote testing before the SARS-CoV-2 outbreak, progress in this field needs to be accelerated. Once social distancing measures are relaxed, it is reasonable to expect a gradual return to normality. However, it is unrealistic to expect our way of life to be wholly unchanged. With the world turning to technologies that facilitate virtual interactions, there are likely to be technical improvements made to these tools, as well as increased availability. The discovery that certain virtual experiences are equally efficient as, or more efficient than, their real-world counterparts may change the way that many of us operate. Considering the widespread effects of the current pandemic and the potential for similar infectious disease pandemics in the future, it is realistic to expect that virtual research will become increasingly popular and, perhaps, even the new norm [5]. Now, more than ever, resources need to be invested in the development of remote research assessments.

Here, we discuss the benefits of conducting remote clinical research, how to determine the suitability of an assessment for remote research, and various approaches to validating such assessments based on where and how frequently data collection occurs. We focus on the validation process of active assessments, including both objective and subjective measures (neuropsychological tests and patient-reported outcomes, respectively). However, the principles outlined here should also apply to the validation process of some passive assessments, such as those designed to detect cigarette smoking [6]. We have included a flowchart (Figure 1) to illustrate the decision-making process. We discuss traditional validation techniques (ie, the comparison of a single assessment between controlled and uncontrolled settings and comparison between two different assessments in the same setting), as well as innovative methods that account for the measurement of constructs across time and location (eg, improving signal-to-noise detection to capture more robust baselines and develop novel phenotypes for improved precision in diagnosis and accuracy in outcomes). We have included case studies to illustrate the breadth of approaches and techniques that may be necessary to consider when designing a validation process for novel assessments. However, it is important to note that these individual pilot studies are only presented here for example purposes and should not be considered comprehensive empirical studies in their entirety.

Figure 1. The decision process for validating digital assessments for remote research.



The Benefits of Conducting Clinical Research Remotely

In a conventional clinical trial, researchers ask patients to complete comprehensive assessments to monitor their symptoms. The assessments can be extensive and may require trained personnel to administer and score them. As the assessments are burdensome for both parties [7], the frequency with which they can be administered is limited and, as a result, they only provide snapshots of treatment efficacy. In other words, assessments may only be administered monthly or even less frequently. However, symptoms can fluctuate from week to week, day to day, and even within a single day. At the time of assessment, a patient's symptoms might improve or be exacerbated by chance because of factors unrelated to treatment efficacy (eg, a stressful event occurring in the morning before an afternoon assessment). Therefore, it can be difficult to ascertain whether changes in patients' symptoms are due to treatment or extraneous factors. Researchers might test patients on the same day of the week or at the same time of the day throughout the trial to account for fluctuations in symptoms. However, this strategy operates on the assumption that symptom changes occur in a predictable manner.

Many psychiatric disorders are characterized by irregular circadian rhythms. Therefore, these patient groups might have, in particular, stochastic fluctuations in cognitive function and mood that cannot be easily predicted [8]. Researchers may attempt to evaluate symptoms during the intervals between assessments by relying on retrospective subjective reports from patients [3]. However, self-report measures can be unreliable indications of actual behavior [9-12], especially when a patient's condition affects their insight or their memory [13,14], impairing their ability to accurately recall past events and symptoms. In addition, there is a limit to how much detail can be recalled. Retrospective subjective reports usually ask patients to reflect on symptom changes over a week or month because recalling hourly or daily changes would be unfeasible [15]. To account for confounders that coincide with test days and avoid relying on biased reports of symptoms, researchers may measure the symptoms of interest before and after an experimental manipulation in the laboratory. This design is assumed to act as a proxy for what might occur in the real world. However, it is not clear how well the trigger and resulting behavior in an artificial setting will translate into those occurring in the real world [16]. An alternative is to conduct the trial remotely, using a phone or wearable device to administer assessments in natural settings at regular intervals or in response to state changes, also known as ecological momentary assessment [2,3]. This methodology reduces user burden while increasing the likelihood of the measurements capturing clinically relevant symptoms when they occur in real time, and such approaches will help to revolutionize clinical trials [3,17].

Determining the Suitability of an Assessment for Remote Research

When developing any assessment tool, there needs to be a trade-off between the length or duration of an assessment and the frequency at which it can be administered. On the one hand, the testing time should be kept to a minimum. Increased testing time can cause participants to tire of the assessments [18], which could decrease the accuracy of responses and compliance, thereby increasing attrition. Similarly, knowing that one will need to complete frequent lengthy assessments during a research study may negatively affect recruitment to that study. On the other hand, there needs to be a sufficient amount of data to maximize the precision of the measurements and ensure their accuracy (ie, effectiveness at detecting the presence or absence of the symptom of interest) [19]. These considerations are especially crucial when the research study requires participants to interrupt their normal routines to complete assessments [20,21] and even more so when conducting clinical research with patients who may have a low threshold for burdensome research procedures because of their symptoms (eg, lack of concentration, fatigue, or motivational fluctuations). Therefore, for the most part, remote assessments should be either *brief and frequent* or *lengthy and infrequent* to be successfully implemented and to reliably capture valid data. However, there are exceptions to this rule, which we discuss herein.

Assessments that require a substantial amount of time to complete cannot be easily incorporated into daily routines or administered frequently without overly inconveniencing users. The inclusion of these assessments in real-world trials is likely to decrease compliance and increase attrition. Therefore, lengthy assessments will need to be abbreviated to be administered at high frequencies.

However, some assessments cannot be abbreviated, such as those that cannot sacrifice items or trials without degrading the assessment's psychometric properties. As long as lengthy assessments can be administered at relatively low frequencies (eg, once a day or once a week) and have some flexibility regarding when they can be completed (within reason), they can be administered remotely (ie, on the web). Administering a lengthy assessment remotely (as opposed to administering it at a testing facility) provides a more naturalistic context for data collection and may reduce confounding factors associated with artificial test settings. It also reduces the intervals between events of interest and subsequent measurements, which may improve symptom recall accuracy. In addition to these benefits, there are temporal and spatial limitations within this context that should be considered. Users will need to find an appropriate space as well as time to engage with the assessment. Therefore, administering lengthy assessments remotely is not a suitable method for capturing clinically relevant data in real time. However, depending on the research question and clinical population, it might be preferable for participants to complete an extensive assessment less frequently and retrospectively, rather than complete a less comprehensive assessment more frequently and in real time (eg, when qualitative data are needed, for the purposes of a clinical diagnostic interview, etc).

Assessments that are brief (or can be abbreviated), track dynamic changes, and are not limited by specific technical requirements are appropriate for high-frequency remote testing. This approach captures high-resolution data that allow for the interpretation of outcomes in relation to time and location. This increased level of detail can be incorporated into the statistical analyses to improve the signal-to-noise ratio. As a result, high-frequency assessments can be used to achieve more representative baselines and develop novel digital phenotypes to improve the precision and accuracy of diagnosis and outcomes [22]. It should be noted that brief assessments do not necessarily have to be administered at high frequencies to be valid measurement tools. For example, the 2-item Patient Health Questionnaire (PHQ) is an abbreviated version of the 9-item PHQ, which is widely used to evaluate depression and demonstrates sufficient diagnostic sensitivity when administered at the same low frequency as the full-length version [23]. Administering the abbreviated version infrequently is likely to be less sensitive than both administering the full-length version infrequently and administering the abbreviated version frequently. However, the latter options may not always be feasible (eg, when the other study procedures are already time consuming and effortful) or sensible (eg, when increasing the response rate or completion rate is key).

Sampling frequency matters—both over- and undersampling can have negative consequences. It may be inefficient to use hourly sampling to capture diurnally varying symptoms [24] or to use time-based sampling to capture symptoms that occur in response to specific (eg, clinically relevant) events [3]. This is particularly true if the frequency and regularity of the relevant event vary considerably among individuals (eg, panic attacks can occur several times per week or a couple of times per year) [25-27]. Sampling more than necessary risks burdening participants, wasting resources, and ultimately degrading data quality (eg, by decreasing compliance) [24]. The sampling frequency should align with the fluctuations of the symptoms of interest as much as possible so that each measurement is informative. This may mean that low-frequency sampling is the most appropriate, circumventing the need for brief assessments. However, compliance may be low if participants are required to self-initiate assessments after the occurrence of relevant events. Alternatively, high-frequency sampling can be used to continuously monitor relevant events (eg, through a watch that passively detects smoking) and trigger an assessment when appropriate [6,28,29]. There are also cases in which the relevant events may occur irregularly or infrequently, but continuous monitoring of the symptoms of interest outside of the event window is useful; for example, measuring positive and negative affect regularly as well as whenever self-harm occurs [30]. Continuous monitoring of symptoms provides a clearer picture of baseline functioning, which can be used to better characterize changes in functioning. Furthermore, continuous monitoring of symptoms of interest, relevant events, or related factors can help identify patterns in behavior that can be used to predict changes, increasing sampling accuracy and reducing participant burden [31,32]. Of course, the feasibility of continuous monitoring will depend on the effort required by the participant to complete the assessment and the capacities and constraints of the specific participant group.

There are also assessments that are appropriate to abbreviate and administer at high frequencies but may not be suitable for remote testing for other reasons (eg, if researcher supervision or a specialist device is required). Furthermore, if the assessment is susceptible to practice effects, mitigating solutions will need to be developed [33]. Some considerations may not be directly related to the assessment itself but to the context in which it is used. For example, an assessment may be designed to evaluate outcomes after a pharmacological challenge and, in this case, whether or not the pharmacological challenge can be delivered remotely needs to be considered. It may be possible, if appropriate precautions are taken, to instruct participants to self-administer certain substances, such as caffeine, alcohol, nicotine, etc, but this would clearly not be possible in other cases (eg, controlled substances).

A Process for Evaluating Tools for Remote Research

As research transitions from operating in testing facilities to the field, it is vital that remote research assessments are developed to a high standard. A remote research assessment needs to be both a valid measure of the construct being evaluated and practical to implement. One of the difficulties in transitioning to remote data collection lies in defining the process for validating remote assessments. For a novel assessment to be valid, it must be reliable and a true measure of the construct of interest. Reliability can be verified by measuring the internal consistency of items or trials or by investigating whether the assessment produces consistent results under similar conditions (ie, test-retest reliability) [34]. To demonstrate internal validity, a reliable assessment is compared with a gold standard (ie, a tool that has been demonstrated to consistently and accurately measure the construct of interest) under controlled conditions to reduce extraneous influences. To demonstrate external validity or generalizability, the assessment is administered and compared across different testing conditions (ie, at different times, in different settings, and in different people) [35,36].

However, this paradigm for evaluating internal validity is not necessarily useful for validating remote research assessments, which are not designed to be administered under controlled conditions. This is not necessarily a limitation because using a traditional validation paradigm may not be ideal when the focus is on real-world behavior, as is the case in applied research. Testing under controlled conditions can introduce temporal and spatial biases into the data. Unlike traditional research assessment, remote research assessment is far more flexible in terms of when and where it can be administered. This increased flexibility in data collection can improve the external validity of the assessment but also means less standardization because assessments are completed without researcher supervision and in contexts that can vary within and across participants. Therefore, the framework for evaluating the *internal* validity of remote research assessments may need to be different from traditional methods that assume spatiotemporal consistency.

How to Validate Low-Frequency Assessments for Remote Research

The internal validity of an assessment includes face, content, criterion, and construct validity [36,37]. Criterion validity is useful to assess when evaluating the construct validity of an assessment, abbreviating an already existing assessment, or planning to use an assessment in a new environment. To assess the criterion validity of any new assessment (either a completely new assessment or an amended version of an already validated assessment), the validity assessment needs to be administered (concurrent validity) or after (predictive validity) an established assessment. An established assessment is one that has already been validated to measure the same construct or a similar constructs (to evaluate criterion and construct validity, respectively) [36]. The outcomes generated by the new assessment need to be compared with those generated by an established assessment. To validate a new low-frequency assessment for data collection in a clinical or laboratory setting, both the new and established assessments need to be administered under standardized conditions at the testing facility to confirm that the assessments are equivalent. To validate a new low-frequency assessment for remote data collection, the new assessment needs to be administered remotely and the resulting outcomes compared with those generated by the established assessment. The established assessment can either be administered in a clinical or laboratory setting or remotely (depending on whether the assessment has already been validated for remote data collection) [38-40]. When validating assessments remotely, the unsupervised and uncontrolled nature of the study environment and the potential for selection bias need to be considered. Table 1 illustrates not only some limitations of remote data collection but also the advantages that remote data collection offers over data collection at testing facilities. The advantages may offset the disadvantages of remote data collection because research suggests that data collected remotely and in-person are comparable [1,39,41-45].

Table 1. Key factors to consider when validating assessments for remote research.

Factors	Limitations	Advantages
Absence of rater or supervision	• The researcher cannot observe participants to determine whether participants are incapacitated, disengaged, or require clarification and intervene if necessary [46].	• Participants may be less influenced by social facilitation or impairment and behave more naturally [47].
No central testing location and testing can occur at unspecified times	• There may be a higher likelihood of distractions during data collection [48]. • The sample might be biased toward individuals with technology and internet access and technology proficiency [1,39,48,49].	• Being outside of the laboratory or clinic may reduce evaluation apprehension and cause participants to behave more naturally [50]. • Depending on the study design, participants may be reporting on behaviors, mood states, etc when and where they naturally occur [3]. • Participation in the study is accessible to individuals who are unwilling or unable to travel to a central testing location or to be tested in person [1,48].
Differences in device, computer hardware, software, processing speed, screen resolution, display characteristics, internet connection, and response input method	• May bias stimulus presentation and response measures, especially reaction time [1,40,45,48,51,52] • Differences in the ownership of certain devices (eg, smartphones) may be patterned by sociodemographic factors [53].	• Having participants use their personal devices to input data may reduce study costs (devices do not need to be purchased and supplied to participants). In addition, the use of a familiar device may improve performance and compliance [54].

Evaluating behavior under controlled (ie, laboratory or clinic) or quasi-controlled (ie, on the web) conditions may suffer from poor external validity because the findings will not necessarily represent natural behavior. External validity refers to the degree to which the measurements generated by an assessment generalize to other people (population validity) and settings (ecological validity) [35] and across time [55]. In field research, behavior can be evaluated at frequent intervals, in natural settings, and in real time. This avoids experimenter and recall biases, increasing the ecological and temporal validity of the research. The population validity of field research is less straightforward (Textbox 1).

Textbox 1. Taking a closer look at the external validity of remote assessments.

Selection biases

- Both remote and in-person studies are subject to selection biases [49,56-60]. Whether a study is conducted remotely or in-person, participants are motivated to take part for a variety of reasons. For example, a common motivation for taking part in in-person studies is financial reward; other reported motivations include the desire to help science and medicine, help other people, learn, and socialize [60]. Differences in participation motivation can have downstream effects such as affecting engagement with the study procedures. In turn, data quality may suffer, resulting in misleading findings. Previous research demonstrates that there are systematic differences in engagement between paid and course-credit participants in in-person studies [59] and between web-based participants looking for paid work (eg, *Amazon Mechanical Turk* users) and those recruited through paid advertisements [56].
- A large proportion of the participants in in-person studies are Western, educated, affluent, and democratic individuals from industrialized countries [61] and primarily students [1,48]. The resulting lack of diversity of the sample can weaken population validity [1,48,61]. For example, there is evidence of systematic differences in data obtained from student samples and the general population [57]. When conducting research with clinical populations, there are additional barriers associated with poor health that can bias trial recruitment and retention [62]. Individuals with the lowest levels of functioning may be the least likely to participate in, or the first to drop out from, clinical trials because participation might be too burdensome [63]. Conversely, when recruitment for clinical trials primarily occurs at health care facilities, individuals who do not visit doctors' offices and hospitals (perhaps those with the highest levels of functioning or those that dislike or fear health care settings) are likely to be underrepresented in the clinical research [64]. Collecting data through remote assessments may increase sample diversity, for instance, by making participation more accessible to nonstudent populations (such as individuals who work during normal operating hours, who have care responsibilities, who live and work far from the university, who are unfamiliar with research, etc). Collecting data through remote assessments can also increase sample diversity for clinical research by making participation more accessible to individuals with varying disease severity and to those reluctant to seek out treatment. Remote methodology also allows individuals who might not otherwise participate in research because of disapproval from family or friends [65] to participate discretely.
- However, although this approach mitigates certain selection pressures, it is likely to induce different selection effects relating to, for instance, internet and device access [49,54,66,67]. Although it is commonly accepted that this may affect generalizability, it may also bias exposure-outcome relationships within the study of interest owing to collider bias. For example, say we enroll participants in a remote study on cognitive performance in which assessment necessitates using an iPhone. It has been demonstrated that ownership of an iPhone is associated with educational attainment, age, and health [68]. If we assume that educational attainment is related to cognitive performance, then any relationship we see between the predictors of iPhone use (eg, age and health) and cognition may be distorted by collider bias [69].
- Instead, researchers may allow participants to complete the assessments on any smartphone to increase the inclusivity of the research. If large variations in responses due to software or hardware differences are anticipated, the analysis may include device type as a covariate to control for this variability. Doing so, however, can again introduce collider bias, where, for example, an association between socioeconomic status and cognitive performance may appear weaker than the true population value [54]. This selection bias poses a risk to generalizability of the findings. Therefore, researchers must carefully consider their recruitment strategy and implement statistical tools such as weighted and sensitivity analyses to avoid and correct for selection biases [58,70].

How to Validate High-Frequency Assessments for Remote Research

Field assessments cannot feasibly be administered in controlled settings at fixed times in an attempt to avoid interference from the outside world. As a result, it can be challenging to empirically evaluate the impact of extraneous factors and thus demonstrate robust internal validity, especially because there can be substantial intraindividual variability in many important symptoms and behaviors. One solution is to exploit the ability of these tools to capture high-resolution data [22]. High intraindividual variability can inflate the sample SD; increasing the number of data points per participant can increase the precision of estimates and improve statistical power [71-74]. However, the feasibility of increased sampling needs to be considered because it can exacerbate practice and fatigue effects [33].

To validate any new high-frequency assessment, the procedure is broadly the same as that for a low-frequency assessment: outcomes from a high-frequency assessment can be compared with those generated by an established low-frequency assessment. As high-frequency assessments are administered under flexible conditions, they exhibit greater external validity at the expense of internal validity. Likewise, because low-frequency assessments are administered under stricter conditions, they exhibit greater internal validity at the expense of external validity (Textbox 1). Therefore, equivalence in the outcomes generated by these complementary measures suggests that high-frequency assessments are likely to possess robust external and internal validity.

However, to make a comparison between two complementary measures, any methodological differences in how the measures are implemented need to be taken into account. When validating an assessment for high-frequency testing, the new assessment is often a brief assessment, whereas the established assessment may be a full-length assessment. When validating a brief assessment against a full-length assessment, there is often a temporal mismatch. A full-length assessment needs to be administered only once to provide meaningful data. However, because a brief assessment is less comprehensive than a full-length assessment, a single data point may be less likely to be informative. Instead, the brief assessment needs to be administered repeatedly across a range of time points, with the resulting multiple data points taken together to provide useful information.

An exception might be when a full-length assessment is not particularly lengthy to begin with; therefore, the abbreviated version is not considerably shorter than the full-length version, and the psychometric properties of the assessment are not drastically affected. The reason for shortening the assessment might be to coadminister it alongside other assessments while keeping the total testing time brief. Alternatively, it might be beneficial to remove a component that may be problematic when delivering assessments remotely. For example, the 8-item PHQ is equivalent to the 9-item PHQ except that it excludes the item on suicidal ideation. It is useful for screening for depression in environments where it would not be feasible to implement safeguarding procedures for participants who indicate suicidal thoughts or intentions [75]. In these cases, the assessment may still be administered at high frequencies, but not because increased sampling is necessary to compensate for a reduction in trials or items. Therefore, in these cases, the outcomes from a single abbreviated assessment would be meaningful and could be directly compared with the outcomes from a single full-length assessment.

When validating a brief assessment against a full-length assessment, there is often a spatial mismatch. Brief assessments are often implemented to facilitate field research, whereas full-length assessments are best suited to more controlled environments where participants can dedicate a substantial amount of time to attend to the assessment (ie, in a clinical or laboratory setting). Therefore, when comparing outcomes between the two it is useful to compare high-frequency measurements collected in the field with low-frequency measurements collected in a controlled environment. For example, a recent study evaluated the feasibility and validity of high-frequency cognitive assessments in patients with schizophrenia. Patients and healthy controls completed a traditional neuropsychological battery at the clinic, followed by an ecological momentary assessment (hosted on a mobile phone) to measure cognitive function remotely for 7 days. Compliance was high, fatigue effects were not observed, and practice effects occurred as a function of study duration, but this relationship was observed for both the patient and control groups. Outcomes for the high-frequency abbreviated assessments correlated considerably with the outcomes from the validated full-length assessments in both patients and controls, demonstrating convergent validity for the high-frequency assessments [76]. However, the full-length assessment, against which the abbreviated assessment is validated, does not necessarily need to be administered at a testing facility. It can instead be administered remotely if it has been validated for remote administration and can feasibly be administered at moderately high frequencies (eg, once a week or once a day). This approach is illustrated in Textbox 2 [77].

Textbox 2. Comparing high-frequency abbreviated assessments with low-frequency full-length assessments.

Aim

- To evaluate the feasibility and validity of high-frequency assessments to capture fluctuations in cognition and mood

Methods

- Ecological momentary assessment was used to measure cognition and mood remotely for 2 weeks in 10 healthy participants in a pilot study.
- Cognitive function was assessed remotely by using the following:
    - Validated full-length assessments from the Cambridge Neuropsychological Test Automated Battery (CANTAB): spatial working memory, rapid visual information processing, attention switching task, and emotion recognition task. The assessments were hosted on a web page and administered after 5 PM each day.
    - An abbreviated assessment (hosted on a smartwatch: Microsoft Band 2) of working memory (A-prime: the ratio of hits [correct detection of an n-back match] to false alarms [response during no match] 2-back task). The assessments were administered once per hour between 9 AM and 7 PM.
- Mood was assessed remotely by the following methods:
    - A validated full-length assessment: positive and negative affect schedule. The assessment was hosted on a web page and administered after 5 PM each day.
    - A brief assessment (hosted on a smartwatch: Microsoft Band 2) of emotional state (through the selection of the participant's current emotion and rated intensity of this emotion) probed immediately after cognitive testing each day.

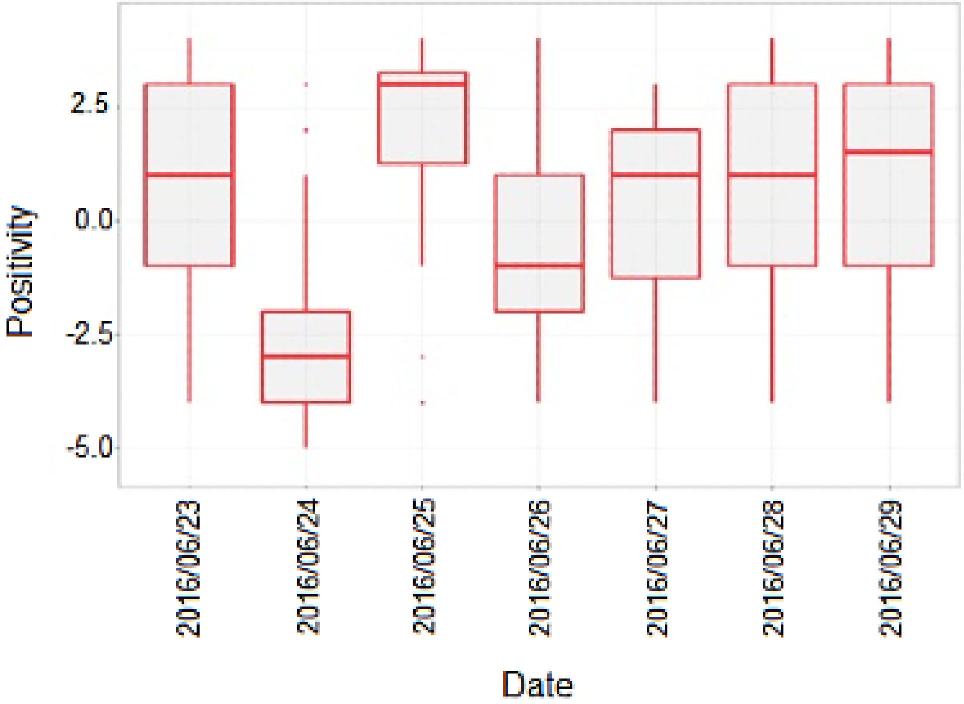
Key findings

- The feasibility of the high-frequency methodology was evaluated by measuring compliance with data collection. The high-frequency 2-back task was completed on 64% (9/14) of the study days, with an average of 3.6 tests completed on those days. More assessments were completed on weekdays than on weekends and outside of commuting hours (9 AM and 6 PM).
- The convergent validity of the high-frequency 2-back task was evaluated by correlating its outcomes with the outcomes from the CANTAB tests. A-prime was significantly correlated with measures of spatial working memory (r=−0.8) and attention switching task (r=−0.45), and moderate but not statistically significant correlations were observed with performance on a measure of sustained attention (rapid visual information processing A-prime r=−0.33). On the high-frequency assessments of mood, this nonclinical sample rated the mood as generally positive and of a low intensity. Participants were first asked to rate their emotion by choosing 1 of 6 canonical emotions (happiness, sadness, disgust, fear, surprise, and anger) and then the intensity of that emotion on a 6-point scale where 6 was the most intense. As negative emotions (sadness, disgust, fear, and anger) were much less frequently reported than positive emotions (happiness and surprise) in this healthy sample, daily intensity reports across positive and negative emotions were aggregated to produce a single scale representing the overall balance of reports of positive or negative emotional intensities over a day. Notably, a reduction in mood positivity was observed on the day of the results of the 2016 United Kingdom European Union membership referendum (June 24, 2016; Figure 2).

Key conclusions

- A full-length assessment allows for comprehensive data collection at a single time point and allows for researchers to exert greater control over the testing environment. Yet, the data might be distorted owing to low-sampling frequency or use of an artificial environment. In this case study, it was feasible to administer the full-length assessments daily in natural settings. However, many full-length assessments might be too long to administer as frequently as once a day [78] or need to be administered at a testing facility [76] (eg, when specialist equipment or a trained administrator is required). The results in this case study demonstrate how extraneous factors (eg, the referendum) can affect outcomes. Outcomes that are measured infrequently are more vulnerable to confounding bias (ie, the outcomes may differ dramatically depending on the day or time when they were measured). Measuring outcomes at high frequencies, instead, allows researchers to detect confounding effects and control for, or investigate, them as appropriate.
- To feasibly measure outcomes at high frequencies, assessments must be brief, which means the comprehensiveness of data collected at a single time point is drastically reduced. Therefore, to ensure that abbreviated assessments are sensitive to what they are intended to measure, they need to be administered more frequently than, say, once per week or once per day. This often requires sampling to occur in real-world environments because it is impractical to collect data in a laboratory or clinic at such high frequencies. Using field assessments comes with its own set of challenges; therefore, study designs should account for times when engagement may be low (eg, weekends or during commuting).
- Although it is unfeasible to validate field assessments under tightly controlled conditions, they can be compared with assessments that have reliably exhibited internal validity in both healthy participants and patient groups, such as the CANTAB tests portrayed in this case study [79-83]. Field assessments benefit from sampling phenomena within natural settings and in real time over extended periods of time. When outcomes generated by the field assessments are comparable with those generated by validated full-length assessments, it is reasonable to assume that the field assessments also exhibit strong internal validity or, at the very least, are sufficiently valid measures owing to robust external validity and extensive sampling.

Figure 2. Daily mood positivity across all participants over a 7-day period.



One disadvantage of validating a high-frequency assessment against a low-frequency assessment is that it seeks equivalence between outcomes captured at different times and in different locations. An alternative approach is to compare 2 assessments that take measurements in equivalent ways (ie, at high frequencies in the field). This approach is only possible when a brief assessment (validated to measure the same or an empirically similar construct of interest) already exists. This approach allows for a new brief assessment to be validated against an established assessment in real time, evaluating construct and criterion validity. As the assessments are completed concurrently, both assessments will be subject to similar influences (eg, common confounding structures). However, this threat to internal validity is likely to be offset by the richer and more granular data produced by high-frequency assessments. It allows for in-depth exploration of interindividual and intraindividual variability, which is key to identifying a signal in a noisy setting. This validation approach is illustrated in Textbox 3. Although one of the strengths of high-frequency testing is increased external validity (Textbox 1), it is worth thinking critically about the degree to which it increases generalizability (population validity) specifically, with special consideration given to recruitment strategy and analytical approaches [58,70].

Textbox 3. Comparing different high-frequency assessments.

Aim

- To evaluate the feasibility and validity of a high-frequency assessment of vigilant attention and explore how reducing task length affects both factors

Methods

- Ecological momentary assessment was used to measure vigilant attention remotely for 2 weeks in 13 healthy participants in a pilot study.
- Vigilant attention (which is sensitive to sleep deprivation) was assessed remotely by using the following:
  - An abbreviated version of the psychomotor vigilant task (PVT), an objective measure of vigilant attention. The PVT measured reaction time after stimulus onset across approximately 50 trials. It was hosted on a mobile phone and was administered up to 2 times per day (morning and afternoon).
  - A subjective, validated measure of sleepiness or alertness, the Karolinska Sleepiness Scale (KSS). The KSS consists of a single rating of sleepiness or alertness on a 9-point scale. The KSS (hosted on a mobile phone) was administered immediately after each administration of the PVT.

Key findings

- Compliance was poor; of the 13 participants who took part, 10 completed at least 50% (14/28) of the assessments. However, it should be noted that the PVT and KSS measures were administered as part of a longer battery (approximately 10 min), which many of the participants felt was burdensome. Therefore, it is probable that compliance would have been higher if the PVT and KSS measures were administered on their own. Overall, mean PVT reaction times were correlated moderately with KSS scores (r=0.37; 95% CI 0.25-0.48). As the outcome measures were captured in real time and sampled frequently, they were not influenced by retrospective recall bias and were less susceptible to coincidental factors. In addition, the granularity of the data allowed for in-depth analysis of how compliance, task performance, and task sensitivity changed as a result of repeated assessments, the time of day, the day of the week, task length, and individual differences.
- For example, to assess if the PVT's sensitivity to sleep deprivation changed as a result of task length, the association between PVT reaction times and KSS scores across all time points for the full 50 trials can be compared with the first 45, 40, and so on, trials. As there are multiple observations for each individual, a mixed model can be used to account for the dependency of observations, where observations (level 1) are nested within participants (level 2). Below, we have plotted the proportion of variance explained by the model based on the number of trials included in the analysis (Figure 3).
- The plot illustrates that the amount of variation in PVT reaction time explained by KSS scores does not change when fewer trials are included in the analysis. However, this plot depicts aggregate data across all participants. Owing to the granular nature of the data, several data points exist for each measure for each participant. Therefore, it is possible to calculate an intraindividual correlation between 2 variables (such as PVT reaction times and KSS scores) for each participant to allow for the interpretation of interindividual and intraindividual variability. Below, we show the association between PVT reaction times and KSS scores across all time points for each participant with more than 50% (14/28) compliance (n=10) based on the number of trials included in the analysis (Figure 4).
- In addition, owing to the granular nature of the data, sources of variability such as temporal influences on PVT performance (eg, time of day, repetition of assessments) can be explored and accounted for in the modeling. Below, we show how mean PVT reaction time (averaged across all participants) varies based on time of day (morning vs evening assessments) and over the 14-day testing period (Figure 5).

Key conclusions

- The rich, granular data produced by the high-frequency assessments allow for in-depth exploration of interindividual and intraindividual variability. Furthermore, sources of random variability can be accounted for in the analysis, increasing the signal-to-noise ratio.

Figure 3. The proportion of variance explained by the model (conditional R^2) across all participants based on the number of trials that were included in the analysis.

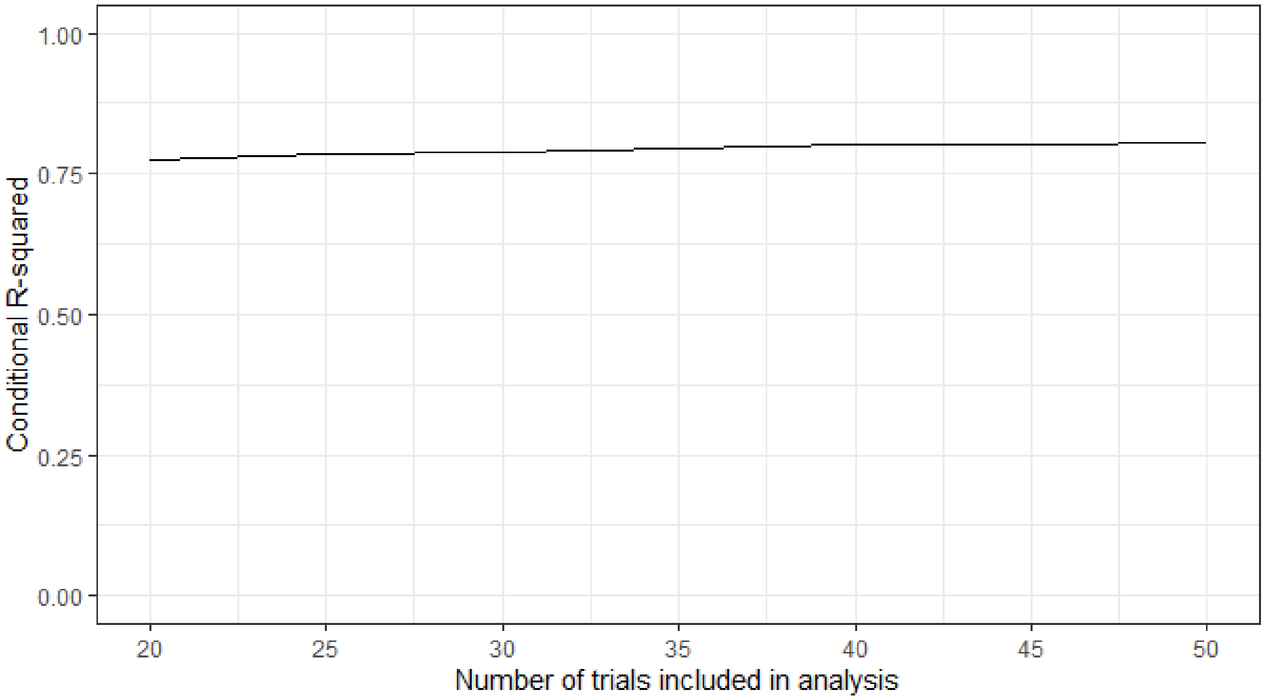


Figure 4. The correlation (Pearson *r*) between psychomotor vigilance task reaction times and Karolinska Sleepiness Scale scores across all time points within participants. Only participants with more than 50% compliance (ie, completed at least 14 of the 28 possible assessments) are included (n=10).

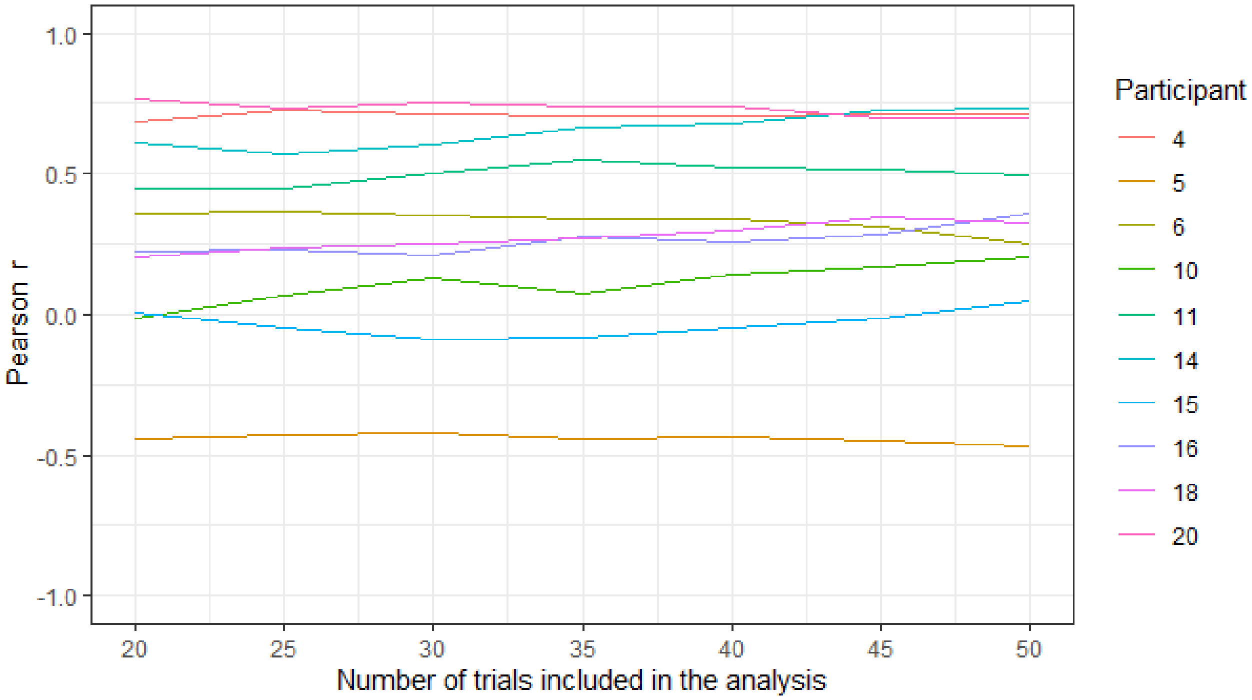
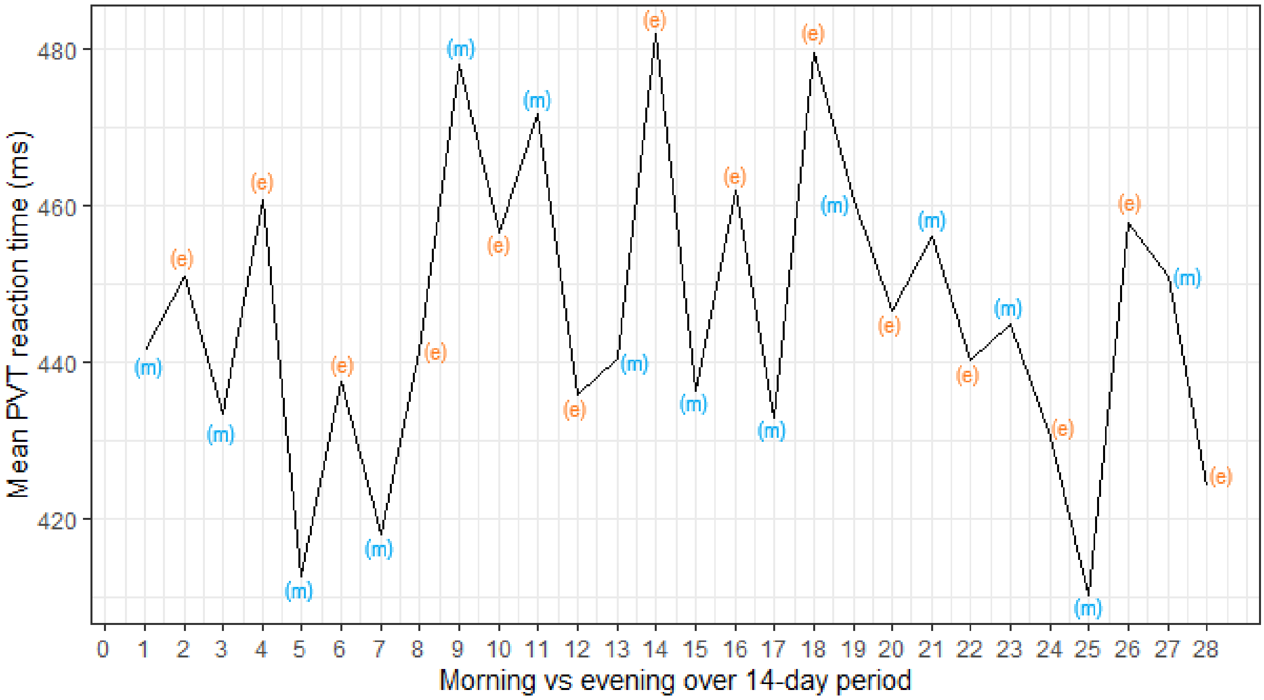


Figure 5. Mean psychomotor vigilance task reaction time (across all participants) as a function of the time of day (morning vs evening) and study day (day 1 to day 14). PVT: psychomotor vigilance task.



Longitudinal data sets, such as those presented in Textbox 2 and Textbox 3, can be analyzed using mixed-effect models, which allow for both fixed and random effects to be included in the modeling. The benefits of using mixed-effects models are that they can tolerate missing data and evaluate changes over time. Furthermore, changes over time can be explored with respect to how each individual changes over time and how this differs among individuals [84,85]. New approaches to mixed-effects modeling are being developed that allow for close investigation of within-individual volatility. This approach is illustrated in Textbox 4.

Textbox 4. Deriving novel phenotypes using fine-grained repeated observations.

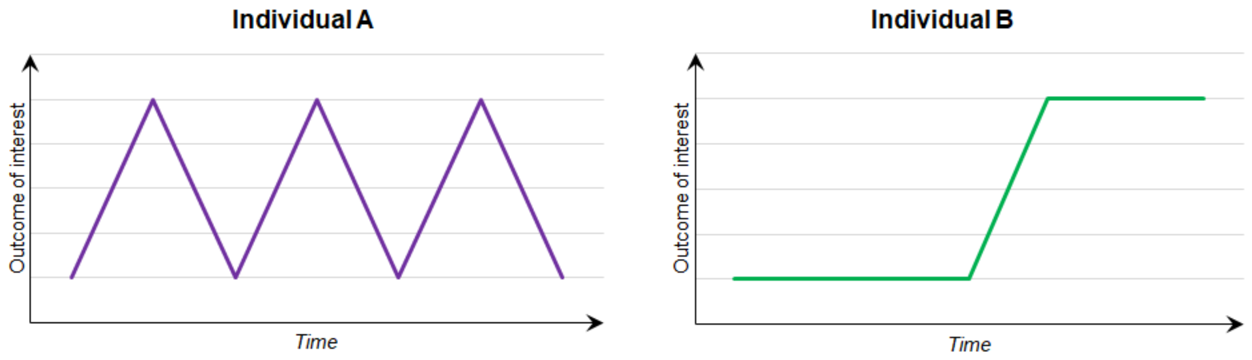
Measuring response volatility

- Fine-grained temporal data such as those offered by the more rapidly reflexive nature of remote research assessments allow researchers to test hypotheses that could not be tested with coarser temporal coverage. For instance, ecological momentary assessment studies can be deployed more rapidly than traditional surveys owing to their electronic distribution—meaning they also allow researchers to ask more reflexive questions, for instance, about the mental health impact of rapidly evolving events such as COVID-19 lockdown policies [86].
- There are further benefits to the use of remote research assessments in the generation of higher-order individual-level characteristics. Repeated measures collected from an individual over time allow for inference about not only the nature of a static response characteristic, but also the within-individual heterogeneity within the response of interest. This is clearly of particular interest in psychological research if it is hypothesized that the variability itself is of substantive interest. For instance, work on borderline personality disorder can require the characterization of affective dysregulation through response volatility [87]. Similarly, studies have investigated associations of affective volatility with mental health and alcohol consumption mediated by mean positive affect in mothers [88].
- Broadly, further increases in temporal granularity allow more complex parameterizations of volatility. We may start with something relatively commonplace such as SD or variance. However, consider the example below: intuitively we can see that individuals A and B have different levels of *volatility*; yet, simply considering the SD or variance of measures will yield identical values for both participants (Figure 6).
- If the researcher wants to distinguish between these individuals, then variance or SD is clearly insufficient. We must include further consideration of, say, autocorrelation or stability [89], or even something more bespoke. For instance, researchers monitoring continuous blood glucose levels from wearable technologies derived a measure of "variability from one moment to the next," operationalized as the length of the line on a graph between 2 adjacent time points [90].
- The fine-grained data afforded to researchers by remote research assessments allow the generation of more complex research questions. For example, take the data presented in Textbox 3. Novel mixed-effect models could be specified to further explore the association between sleep quality and alertness. This would allow for analysis of not only whether sleep quality informs mean levels of alertness, but also whether the variability of the responses of a given individual are predicted by their indicated sleep quality.

Key conclusions

- Fine grained, repeated temporal measures allow researchers to derive novel phenotypes from repeated observations of a given outcome. Extracting and modelling higher order observational phenomena will, in turn, enable better understanding of underlying, within-individual processes underpinning effects in traditional observational enquiry.

Figure 6. Within-individual repeated observations of an outcome of interest with identical means and SDs but different volatility.



Conclusions

Remote research assessments can be used to study cognition and behavior in unconventional and innovative ways while carefully adhering to established research principles. As a result, the use and further development of these assessments will reshape psychological and clinical research in the near future.

These tools are not without their own set of unique challenges and require the careful consideration of the optimal approach, particularly approaches for increasing generalizability, for any given research question. This presents an opportunity for discoveries that, without creative thinking, technological advancements, and flexibility, might otherwise have remained undiscovered. There is always room to improve research tools, and it is vital that the methods to evaluate these tools keep pace.

Acknowledgments

This project was hosted by the Cambridge Cognition Research Hub at the University of Bristol, a collaborative partnership between researchers at Cambridge Cognition Ltd and the University of Bristol. This project falls within the scope of the established collaboration, which is recognized by the University of Bristol through a contract that includes agreed intellectual property terms. All authors contributed to the conceptualization of this work. JF wrote the original draft, and GJG, CS, NC, FKC, JHB, and MRM contributed to the critical revision of this paper. All authors approved the final version of the manuscript for publication.

Conflicts of Interest

MRM provides consultancy to Cambridge Cognition Ltd and is a codirector of Jericoe Ltd, which produces software for the assessment and modification of emotion recognition. JF's post at the University of Bristol is funded by Cambridge Cognition Ltd. CS, NC, NT, JD, FC, and FKC are employees of Cambridge Cognition Ltd. JHB is an employee and shareholder of Cambridge Cognition Ltd.

References

1. Woods AT, Velasco C, Levitan CA, Wan X, Spence C. Conducting perception research over the internet: a tutorial review. PeerJ 2015;3:e1058 [FREE Full text] [doi: 10.7717/peerj.1058] [Medline: 26244107]
2. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
3. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
4. Rajagopalan A, Shah P, Zhang MW, Ho RC. Digital platforms in the assessment and monitoring of patients with bipolar disorder. Brain Sci 2017 Nov 12;7(11):150 [FREE Full text] [doi: 10.3390/brainsci7110150] [Medline: 29137156]
5. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 May 05:1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]
6. Skinner AL, Stone CJ, Doughty H, Munafò MR. StopWatch: the preliminary evaluation of a smartwatch-based system for passive detection of cigarette smoking. Nicotine Tob Res 2019 Jan 04;21(2):257-261 [FREE Full text] [doi: 10.1093/ntr/nty008] [Medline: 29373720]
7. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Forum on Drug Discovery, Development, and Translation, Alper J, Khandekar E, et al. Challenges and opportunities: proceedings of a workshop. The National Academies Press, Washington DC 2019:25502. [doi: 10.17226/25502] [Medline: 31334937]

8. Karatsoreos IN. Links between circadian rhythms and psychiatric disease. Front Behav Neurosci 2014;8:162 [FREE Full text] [doi: 10.3389/fnbeh.2014.00162] [Medline: 24834040]
9. Antikainen R, Hänninen T, Honkalampi K, Hintikka J, Koivumaa-Honkanen H, Tanskanen A, et al. Mood improvement reduces memory complaints in depressed patients. Eur Arch Psychiatry Clin Neurosci 2001;251(1):6-11. [Medline: 11315519]
10. Fredrickson BL. Extracting meaning from past affective experiences: the importance of peaks, ends, and specific emotions. Cogn Emot 2000 Jul;14(4):577-606. [doi: 10.1080/026999300402808]
11. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor GS, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
12. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. Pain 1996;66(1):3-8. [doi: 10.1016/0304-3959(96)02994-6] [Medline: 8857625]
13. Ebner-Priemer U, Kuo J, Welch S, Thielgen T, Witte S, Bohus M, et al. A valence-dependent group-specific recall bias of retrospective self-reports: a study of borderline personality disorder in everyday life. J Nerv Ment Dis 2006;194(10):774-779. [doi: 10.1097/01.nmd.0000239900.46595.72] [Medline: 17041290]
14. Fyock CA, Hampstead BM. Comparing the relationship between subjective memory complaints, objective memory performance, and medial temporal lobe volumes in patients with mild cognitive impairment. Alzheimers Dement (Amst) 2015;1(2):242-248 [FREE Full text] [doi: 10.1016/j.dadm.2015.03.002] [Medline: 26191540]
15. Solhan MB, Trull TJ, Jahng S, Wood PK. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychol Assess 2009;21(3):425-436 [FREE Full text] [doi: 10.1037/a0016869] [Medline: 19719353]
16. Holleman G, Hooge I, Kemner C, Hessels R. The 'Real-World Approach' and its problems: a critique of the term ecological validity. Front Psychol 2020;11:721 [FREE Full text] [doi: 10.3389/fpsyg.2020.00721] [Medline: 32425850]
17. Gold M, Amatniek J, Carrillo MC, Cedarbaum JM, Hendrix JA, Miller BB, et al. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. Alzheimers Dement (N Y) 2018;4(1):234-242 [FREE Full text] [doi: 10.1016/j.trci.2018.04.003] [Medline: 29955666]
18. Ackerman PL, Kanfer R. Test length and cognitive fatigue: an empirical examination of effects on performance and test-taker reactions. J Exp Psychol Appl 2009 Jun;15(2):163-181. [doi: 10.1037/a0015719] [Medline: 19586255]
19. Maxwell SE, Kelley K, Rausch JR. Sample size planning for statistical power and accuracy in parameter estimation. Annu Rev Psychol 2008 Jan;59(1):537-563. [doi: 10.1146/annurev.psych.59.103006.093735] [Medline: 17937603]
20. Edwards P, Roberts I, Sandercock P, Frost C. Follow-up by mail in clinical trials: does questionnaire length matter? Control Clin Trials 2004 Feb;25(1):31-52. [doi: 10.1016/j.cct.2003.08.013] [Medline: 14980747]
21. Edwards P. Questionnaires in clinical trials: guidelines for optimal design and administration. Trials 2010;11:2 [FREE Full text] [doi: 10.1186/1745-6215-11-2] [Medline: 20064225]
22. Cohen AS, Schwartz E, Le T, Cowan T, Cox C, Tucker R, et al. Validating digital phenotyping technologies for clinical use: the critical importance of "resolution". World Psychiatry 2020;19(1):114-115 [FREE Full text] [doi: 10.1002/wps.20703] [Medline: 31922662]
23. Manea L, Gilbody S, Hewitt C, North A, Plummer F, Richardson R, et al. Identifying depression with the PHQ-2: a diagnostic meta-analysis. J Affect Disord 2016;203:382-395. [doi: 10.1016/j.jad.2016.06.003] [Medline: 27371907]
24. Ebner-Priemer UW, Sawitzki G. Ambulatory assessment of affective instability in borderline personality disorder. Eur J Psychol Assess 2007;23(4):238-247 [FREE Full text] [doi: 10.1027/1015-5759.23.4.238]
25. Essau C, Conradt J, Petermann F. Frequency of panic attacks and panic disorder in adolescents. Depress Anxiety 1999;9(1):19-26. [Medline: 9989346]
26. Fyer A, Katon W, Hollifield M, Rassnick H, Mannuzza S, Chapman T, et al. The DSM-IV panic disorder field trial: panic attack frequency and functional disability. Anxiety 1996;2(4):157-166. [doi: 10.1002/(sici)1522-7154(1996)2:4<157::aid-anxi1>3.0.co;2-l]
27. Norton GR, Harrison B, Hauch J, Rhodes L. Characteristics of people with infrequent panic attacks. Journal of Abnormal Psychology 1985;94(2):216-221. [doi: 10.1037/0021-843X.94.2.216]
28. Reichert M, Tost H, Reinhard I, Schlotz W, Zipf A, Salize H, et al. Exercise versus nonexercise activity: e-diaries unravel distinct effects on mood. Med Sci Sports Exerc 2017;49(4):763-773. [doi: 10.1249/MSS.0000000000001149] [Medline: 27824691]
29. Tost H, Reichert M, Braun U, Reinhard I, Peters R, Lautenbach S, et al. Neural correlates of individual differences in affective benefit of real-life urban green space exposure. Nat Neurosci 2019;22(9):1389-1393. [doi: 10.1038/s41593-019-0451-y] [Medline: 31358990]
30. Armey MF, Crowther JH, Miller IW. Changes in ecological momentary assessment reported affect associated with episodes of nonsuicidal self-injury. Behav Ther 2011 Dec;42(4):579-588. [doi: 10.1016/j.beth.2011.01.002] [Medline: 22035987]
31. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: Implications for JITAI optimization. Health Psychol 2015;34S(Suppl):1261-1267 [FREE Full text] [doi: 10.1037/hea0000304] [Medline: 26651467]

32. McClernon FJ, Roy CR. I am your smartphone, and I know you are about to smoke: the application of mobile sensing and computing approaches to smoking research and treatment. Nicotine Tob Res 2013;15(10):1651-1654 [FREE Full text] [doi: 10.1093/ntr/ntt054] [Medline: 23703731]
33. Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. BMC Neurosci 2010;11:118. [doi: 10.1186/1471-2202-11-118] [Medline: 20846444]
34. Green SB, Yang Y, Alt M, Brinkley S, Gray S, Hogan T, et al. Use of internal consistency coefficients for estimating reliability of experimental task scores. Psychon Bull Rev 2016;23(3):750-763 [FREE Full text] [doi: 10.3758/s13423-015-0968-3] [Medline: 26546100]
35. Onwuegbuzie AJ. Expanding the framework of internal and external validity in quantitative research. Paper presented at the Annual Meeting of the Association for the Advancement of Educational Research (AAER) (Ponte Vedra, FL, November 2000). 2000. URL: https://eric.ed.gov/?id=ED448205 [accessed 2021-05-24]
36. Parrott A. Performance tests in human psychopharmacology (2): content validity, criterion validity, and face validity. Hum Psychopharmacol Clin Exp 1991 Jun;6(2):91-98. [doi: 10.1002/hup.470060203]
37. Parrott A. Performance tests in human psychopharmacology (3): construct validity and test interpretation. Hum Psychopharmacol Clin Exp 1991;6(3):197-207. [doi: 10.1002/hup.470060303]
38. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, ISPOR ePRO Task Force. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009;12(4):419-429 [FREE Full text] [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
39. Lumsden J, Skinner A, Woods A, Lawrence N, Munafò M. The effects of gamelike features and test location on cognitive test performance and participant enjoyment. PeerJ 2016;4:e2184 [FREE Full text] [doi: 10.7717/peerj.2184] [Medline: 27441120]
40. Schreiner M, Reiss S, Schweizer K. Method effects on assessing equivalence of online and offline administration of a cognitive measure: the exchange test. Int J Internet Sci. 2014. URL: https://www.ijis.net/ijis9_1/ijis9_1_schreiner_et_al_pre.html [accessed 2021-05-24]
41. Assmann KE, Bailet M, Lecoffre AC, Galan P, Hercberg S, Amieva H, et al. Comparison between a self-administered and supervised version of a web-based cognitive test battery: results from the nutrinet-santé cohort study. J Med Internet Res 2016;18(4):e68 [FREE Full text] [doi: 10.2196/jmir.4862] [Medline: 27049114]
42. Cromer JA, Harel BT, Yu K, Valadka JS, Brunwin JW, Crawford CD, et al. Comparison of cognitive performance on the cogstate brief battery when taken in-clinic, in-group, and unsupervised. Clin Neuropsychol 2015;29(4):542-558. [doi: 10.1080/13854046.2015.1054437] [Medline: 26165425]
43. Feenstra HE, Murre JM, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam Cognition Scan. J Clin Exp Neuropsychol 2018;40(3):253-273. [doi: 10.1080/13803395.2017.1339017] [Medline: 28671504]
44. Silverstein S, Berten S, Olson P, Paul R, Willams LM, Cooper N, et al. Development and validation of a World-Wide-Web-based neurocognitive assessment battery: WebNeuro. Behav Res Methods 2007;39(4):940-949. [doi: 10.3758/bf03192989] [Medline: 18183911]
45. Backx R, Skirrow C, Dente P, Barnett JH, Cormack FK. Comparing web-based and lab-based cognitive assessment using the cambridge neuropsychological test automated battery: a within-subjects counterbalanced study. J Med Internet Res 2020;22(8):e16792 [FREE Full text] [doi: 10.2196/16792] [Medline: 32749999]
46. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Arch Clin Neuropsychol 2012;27(3):362-373 [FREE Full text] [doi: 10.1093/arclin/acs027] [Medline: 22382386]
47. Yantz C, McCaffrey R. Social facilitation effect of examiner attention or inattention to computer-administered neuropsychological tests: first sign that the examiner may affect results. Clin Neuropsychol 2007;21(4):663-671. [doi: 10.1080/13854040600788158] [Medline: 17613984]
48. Skitka LJ, Sargis EG. The internet as psychological laboratory. Annu Rev Psychol 2006;57(1):529-555. [doi: 10.1146/annurev.psych.57.102904.190048] [Medline: 16318606]
49. Blom AG, Herzing JM, Cornesse C, Sakshaug JW, Krieger U, Bossert D. Does the recruitment of offline households increase the sample representativeness of probability-based online panels? Evidence from the German internet panel. Soc Sci Comput Rev 2016;35(4):498-520. [doi: 10.1177/0894439316651584]
50. Greenwood JD. On the relation between laboratory experiments and social behaviour: causal explanation and generalization. J Theory of Soc Behav 1982;12(3):225-250. [doi: 10.1111/j.1468-5914.1982.tb00449.x]
51. Crump MJC, McDonnell JV, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One 2013;8(3):e57410 [FREE Full text] [doi: 10.1371/journal.pone.0057410] [Medline: 23516406]
52. Parsons TD, McMahan T, Kane R. Practice parameters facilitating adoption of advanced technologies for enhancing neuropsychological assessment paradigms. Clin Neuropsychol 2018;32(1):16-41. [doi: 10.1080/13854046.2017.1337932] [Medline: 28590154]

53. Demographics of mobile devices ownership in the United States. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-06-03]
54. Germine L, Reinecke K, Chaytor NS. Digital neuropsychology: challenges and opportunities at the intersection of science and software. Clin Neuropsychol 2019;33(2):271-286. [doi: 10.1080/13854046.2018.1535662] [Medline: 30614374]
55. Munger K. The limited value of non-replicable field experiments in contexts with low temporal validity. Soc Media + Soc 2019;5(3):A. [doi: 10.1177/2056305119859294]
56. Antoun C, Zhang C, Conrad FG, Schober MF. Comparisons of online recruitment strategies for convenience samples. Field Methods 2015;28(3):231-246. [doi: 10.1177/1525822X15603149]
57. Hanel PH, Vione KC. Do student samples provide an accurate estimate of the general public? PLoS One 2016;11(12):e0168354 [FREE Full text] [doi: 10.1371/journal.pone.0168354] [Medline: 28002494]
58. Munafò MR, Tilling K, Taylor A, Evans D, Smith GD. Collider scope: when selection bias can substantially influence observed associations. Int J Epidemiol 2018;47(1):226-235 [FREE Full text] [doi: 10.1093/ije/dyx206] [Medline: 29040562]
59. Nicholls ME, Loveless KM, Thomas NA, Loetscher T, Churches O. Some participants may be better than others: sustained attention and motivation are higher early in semester. Q J Exp Psychol (Hove) 2015;68(1):10-18. [doi: 10.1080/17470218.2014.925481] [Medline: 24842155]
60. Stunkel L, Grady C. More than the money: a review of the literature examining healthy volunteer motivations. Contemp Clin Trials 2011;32(3):342-352 [FREE Full text] [doi: 10.1016/j.cct.2010.12.003] [Medline: 21146635]
61. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci 2010;33(2-3):61-83. [doi: 10.1017/s0140525x0999152x]
62. Gul R, Ali P. Clinical trials: the challenge of recruitment and retention of participants. J Clin Nurs 2010;19(1-2):227-233. [doi: 10.1111/j.1365-2702.2009.03041.x] [Medline: 20500260]
63. Naidoo N, Nguyen VT, Ravaud P, Young B, Amiel P, Schanté D, et al. The research burden of randomized controlled trial participation: a systematic thematic synthesis of qualitative evidence. BMC Med 2020;18(1):6 [FREE Full text] [doi: 10.1186/s12916-019-1476-5] [Medline: 31955710]
64. Joshi V. Transparency in recruiting patients for clinical trials. Perspect Clin Res 2013 Oct;4(4):239-240 [FREE Full text] [doi: 10.4103/2229-3485.120175] [Medline: 24312894]
65. Engaging for increased research participation: public and healthcare professionals' perceptions. University Hospital Southampton NHS Foundation Trust. 2014. URL: https://www.uhs.nhs.uk/Media/Southampton-Clinical-Research/Marketresearch/Engaging-for-increased-research-participation-full-report-v2.pdf [accessed 2021-05-24]
66. Bol N, Helberger N, Weert JC. Differences in mobile health app use: a source of new digital inequalities? Infor Soc 2018;34(3):183-193. [doi: 10.1080/01972243.2018.1438550] [Medline: 26281194]
67. Rich E, Miah A, Lewis S. Is digital health care more equitable? The framing of health inequalities within England's digital health policy 2010-2017. Sociol Health Illn 2019;41 Suppl 1(S1):31-49 [FREE Full text] [doi: 10.1111/1467-9566.12980] [Medline: 31599987]
68. 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 2021-06-03]
69. Rohrer JM. Thinking clearly about correlations and causation: graphical causal models for observational data. Adv Methods Pract Psychol Sci 2017;1(1):27-42. [doi: 10.31234/osf.io/t3qub]
70. Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat Commun 2020;11(1):1-12 [FREE Full text] [doi: 10.1038/s41467-020-19478-2] [Medline: 33184277]
71. Baker DH, Vilidaite G, Lygo FA, Smith AK, Flack TR, Gouws AD, et al. Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. Psychol Methods 2020:1-20. [doi: 10.1037/met0000337] [Medline: 32673043]
72. Phillips G, Jiang T. Measurement error and equating error in power analysis. Pract Assess Res Eval 2016;21:9. [doi: 10.7275/0snn-zm67]
73. Rouder JN, Morey RD, Speckman PL, Province JM. Default Bayes factors for ANOVA designs. J Math Psychol 2012 Oct;56(5):356-374. [doi: 10.1016/j.jmp.2012.08.001]
74. Smith P, Little D. Small is beautiful: in defense of the small-N design. Psychon Bull Rev 2018;25(6):2083-2101 [FREE Full text] [doi: 10.3758/s13423-018-1451-8] [Medline: 29557067]
75. Shin C, Lee S, Han K, Yoon H, Han C. Comparison of the usefulness of the PHQ-8 and PHQ-9 for screening for major depressive disorder: analysis of psychiatric outpatient data. Psychiatry Investig 2019;16(4):300-305. [doi: 10.30773/pi.2019.02.01]
76. 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 Ther Comport et Cogn 2018;28(4):204-213. [doi: 10.1016/j.jtcc.2018.02.002]

77. Cormack FK, Taptiklis N, Barnett JH, King J, Fenhert B. TD-P-017: High-frequency monitoring of cognition, mood and behaviour using commercially available wearable devices. Alzheimer's Dement 2016;12:P159. [doi: 10.1016/j.jalz.2016.06.263]
78. Cormack F, McCue M, Taptiklis N, Skirrow C, Glazer E, Panagopoulos E, et al. Wearable technology for high-frequency cognitive and mood assessment in major depressive disorder: longitudinal observational study. JMIR Ment Health 2019;6(11):e12814 [FREE Full text] [doi: 10.2196/12814] [Medline: 31738172]
79. Talebi M, Majdi A, Kamari F, Sadigh-Eteghad S. The Cambridge Neuropsychological Test Automated Battery (CANTAB) versus the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS) for the assessment of cognitive function in patients with multiple sclerosis. Mult Scler Relat Disord 2020;43:102172. [doi: 10.1016/j.msard.2020.102172] [Medline: 32442887]
80. Bland A, Roiser J, Mehta M, Schei T, Boland H, Campbell-Meiklejohn D, et al. EMOTICOM: A neuropsychological test battery to evaluate emotion, motivation, impulsivity, and social cognition. Front Behav Neurosci 2016;10:25 [FREE Full text] [doi: 10.3389/fnbeh.2016.00025] [Medline: 26941628]
81. Lowe C, Rabbitt P. Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. Cambridge Neuropsychological Test Automated Battery. International Study of Post-Operative Cognitive Dysfunction. Neuropsychologia 1998;36(9):915-923. [Medline: 9740364]
82. Torgersen J, Flaatten H, Engelsen BA, Gramstad A. Clinical validation of cambridge neuropsychological test automated battery in a norwegian epilepsy population. J Behav Brain Sci 2012;02(01):108-116. [doi: 10.4236/jbbs.2012.21013]
83. Smith PJ, Need AC, Cirulli ET, Chiba-Falek O, Attix DK. A comparison of the Cambridge Automated Neuropsychological Test Battery (CANTAB) with "traditional" neuropsychological testing instruments. J Clin Exp Neuropsychol 2013;35(3):319-328. [doi: 10.1080/13803395.2013.771618] [Medline: 23444947]
84. Singer J, Willett J. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. 1st Ed. New York: Oxford University Press, Inc; 2003:1-644.
85. Perret C, Kandel S. Taking advantage of between- and within-participant variability? Front Psychol 2014;5:1235 [FREE Full text] [doi: 10.3389/fpsyg.2014.01235] [Medline: 25400611]
86. Fried EI, Papanikolaou F, Epskamp S. Mental health and social contact during the COVID-19 pandemic: an ecological momentary assessment study. PsyArXiv 2020:1-16 [FREE Full text] [doi: 10.31234/osf.io/36xkp]
87. Santangelo PS, Limberger MF, Stiglmayr C, Houben M, Coosemans J, Verleysen G, et al. Analyzing subcomponents of affective dysregulation in borderline personality disorder in comparison to other clinical groups using multiple e-diary datasets. Borderline Personal Disord Emot Dysregul 2016;3(5):1-13 [FREE Full text] [doi: 10.1186/s40479-016-0039-z] [Medline: 27386138]
88. Maher JP, Ra CK, Leventhal AM, Hedeker D, Huh J, Chou C, et al. Mean level of positive affect moderates associations between volatility in positive affect, mental health, and alcohol consumption among mothers. J Abnorm Psychol 2018;127(7):639-649 [FREE Full text] [doi: 10.1037/abn0000374] [Medline: 30221951]
89. Koval P, Brose A, Pe ML, Houben M, Erbas Y, Champagne D, et al. Emotional inertia and external events: the roles of exposure, reactivity, and recovery. Emotion 2015;15(5):625-636. [doi: 10.1037/emo0000059] [Medline: 25844974]
90. Millard L, Patel N, Tilling K, Lewcock M, Flach P, Lawlor D. Software application profile: GLU: a tool for analysing continuously measured glucose in epidemiology. Int J Epidemiol 2020:14. [doi: 10.1101/500256]

Abbreviations

PHQ: Patient Health Questionnaire

Edited by R Kukafka; submitted 25.11.20; peer-reviewed by L Germine, A Klein, D Gard; comments to author 27.01.21; revised version received 23.03.21; accepted 04.05.21; published 18.06.21.

Please cite as:
Ferrar J, Griffith GJ, Skirrow C, Cashdollar N, Taptiklis N, Dobson J, Cree F, Cormack FK, Barnett JH, Munafò MR
Developing Digital Tools for Remote Clinical Research: How to Evaluate the Validity and Practicality of Active Assessments in Field Settings
J Med Internet Res 2021;23(6):e26004
URL: https://www.jmir.org/2021/6/e26004
doi:10.2196/26004
PMID:34142972

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org